Amide Bond Formation in Nonribosomal Peptide Synthesis:

The Formylation and Condensation Domains

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

dem
Fachbereich Chemie
der Philipps-Universität Marburg
vorgelegt von

Georg Schönafinger

aus München

Marburg/Lahn 2007

Vom Fachbereich Chemie

der Philipps-Universität Marburg als Dissertation

am _____ angenommen.

Erstgutachter : Prof. Dr. M. A. Marahiel (Philipps-Universität, Marburg)

Zweitgutachter : Prof. Dr. L.-O. Essen (Philipps-Universität, Marburg)

Tag der Disputation: 20. Dezember 2007

The majority of the work presented here has been published:

Samel, S.A.[†]**, Schoenafinger, G.** [†]**, Knappe, T.A., Marahiel, M.A., Essen, L.O.** 2007. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* 15(7): 781-92.

[†] equal contribution

Schoenafinger, G., Schracke, N., Linne, U., Marahiel, M.A. 2006. Formylation domain: an essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin. *J Am Chem Soc*. 128(23): 7406-7.

Schoenafinger, G., Marahiel, M.A. accepted 2007. Nonribosomal Peptides. *Wiley Encyclopaedia of Chemical Biology*. In press.

To Anke

Summary

Nonribosomal peptides are of outstanding pharmacological interest, since many representatives of this highly diverse class of natural products exhibit therapeutically important activities, such as antibacterial, antitumor and immunosuppressive properties. Understanding their biosynthesis performed by multimodular mega-enzymes, the nonribosomal peptide synthetases (NRPSs), is one of the key determinants in order to be able to reprogram these machineries for the production of novel therapeutics. The central structural motif of all peptides is the peptide (or amide) bond. In this work, two different amide bond forming catalytic entities from NRPSs were studied: The condensation (C) and formylation (F) domains.

Firstly, the N-terminal F domain of LgrA1, belonging to the biosynthetic machinery required for the production of linear gramicidin was biochemically characterized. Using F-A-PCP_{LgrA1} in *in vitro* experiments, its acceptor substrate specificity towards the template-bound branched aliphatic amino acids valine, isoleucine and leucine was identified. From sequence alignments with other formyltransferases, N^{10}-formyl-tetrahydrofolate (N^{10}-fTHF) was expected to serve as formyl donor in these reactions. This molecule was chemoenzymatically produced and successfully used in the assays. Interestingly, its isomer N^5-fTHF was also accepted even though the apparent product formation speed was 18-fold slower. The necessity of a formylated starter unit for the inititation of the nonribosomal biosynthesis was then tested with the dimodular F-A-PCP-C-A-$PCP_{LgrA1-2}$ enzyme. It was shown that no dipeptide was produced, unless a formyl donor substrate was provided – in which case the formylated dipeptide could be detected. Obviously, the N-terminal formylation of linear gramicidin is critical for its bioactivity, where it functions as an ion channel in a head-to-head dimeric complex that is able to penetrate bacterial cell membranes.

Secondly, the bidomain enzyme PCP-$C_{TycC5-6}$ was used as a model system for C domain studies. Its previously discovered ability to cyclize the PCP_{TycC5}-bound hexapeptide DPhe-Pro-Phe-DPhe-Asn-Gln was further investigated. The head-to-tail connectivity of the cyclic product was proven by MS^n-spectrometry, and the substrate specificity for this reaction was probed in the context of six other oligopeptide substrates, three of which were accepted. Mutational studies were furthermore carried out to scrutinize previously suggested models for the C domain catalyzed reaction. According to one model, a conserved histidine residue of the C domain is involved as a base in the catalytic process. Even though the according alanine and valine mutations produced here led to well-folded soluble proteins, their cyclization activity was found abolished.

Crystallization screens with the wild-type *apo*-PCP-$C_{TycC5-6}$ enzyme afforded one promising condition which was further optimized in collaboration with the crystallographic group of Prof. Dr. Essen (Marburg). Thus, for the first time, the structure of a bidomain enzyme from nonribosomal peptide synthetases was solved – shedding light on so far unknown aspects of inter-domain communication. The relative arrangement of both domains was interpreted as a state in which the *apo*-PCP domain seeks interaction with either a different nonribosomal domain or a phosphopantetheine transferase. As expected, the highly variant so-called linker region that connects both entities was found unstructured, yet weakly interacting with both domains' surfaces. Interestingly, a buffer-derived sulfate ion was seen in direct proximity to the proposed active site of the C domain. It is hypothesized in this work that this tetrahedral anion resembles the transition state of the amide bond forming reaction. Consequently, the transition state would be stabilized by the electrostatic environment of the C domain rather than by chemical base catalysis.

Zusammenfassung

Nichtribosomale Peptide gelten als pharmakologisch hochinteressante Stoffklasse, da zahlreiche ihrer vielfältigen Vertreter wichtige Bioaktivitäten, wie beispielsweise antibiotische, antitumorale oder immunosuppressive Eigenschaften aufweisen. Das Verständnis ihrer Biosynthesen, welche von multimodularen Mega-Enzymen, den nichtribosomalen Peptidsynthetasen (NRPS), bewirkt werden, stellt eine Schlüsselvoraussetzung dafür dar, diese Synthesemaschinerien zur Produktion neuer Therapeutika nutzen zu können. Das zentrale Strukturmotiv aller Peptide ist die Peptid- (oder Amid-) Bindung. In der vorliegenden Arbeit wurden daher zwei verschiedene katalytische Einheiten aus NRPS untersucht, welche die Ausbildung von Amidbindungen katalysieren: Die Kondensations- (C-) und die Formylierungs- (F-) Domäne.

Zum einen wurde die N-terminale F-Domäne aus LgrA1 biochemisch charakterisiert. Sie ist Bestandteil des NRPS-Systems, das für die Biosynthese des linearen Gramicidins notwendig ist. Durch *in-vitro*-Experimente mit dem rekombinanten Enzym F-A-PCP_{LgrA1}, konnte dessen Akzeptor-Substratspezifität für die PCP-gebundenen verzweigt-aliphatischen Aminosäuren Valin, Isoleucin und Leucin aufgedeckt werden. Aufgrund von Sequenzvergleichen mit anderen Formyltransferasen wurde erwartet, dass auch in diesem Falle N^{10}-Formyltetrahydrofolat (N^{10}-fTHF) als Formylgruppendonor dient. Dieses wurde auf chemoenzymatischem Wege hergestellt und erfolgreich in den Formylierungsassays eingesetzt. Interessanterweise, konnte auch dessen Isomer, das N^5-fTHF, eingesetzt werden – obgleich hierbei eine um den Faktor 18 langsamere Produktbildung beobachtet wurde.

Anschließend wurde die Notwendigkeit eines formylierten Startbausteins für die Initiation der nichtribosomalen Biosynthese mit Hilfe des dimodularen Konstrukts F-A-PCP-C-A-$PCP_{LgrA1-2}$ *in vitro* untersucht. Es stellte sich heraus, dass in Abwesenheit des Formyldonors kein Dipeptid erzeugt wird, wohingegen in dessen Gegenwart N_α-formyliertes Dipeptid nachgewiesen werden konnte. Diese Beobachtung legt es nahe, dass die Formylierung am N-Terminus des linearen Gramicidins für dessen Bioaktivität als dimerer Ionenkanal, der sich in die bakterielle Zellmembran einlagert, von großer Bedeutung ist.

Zum anderen wurde das bidomänale Enzym PCP-$C_{TycC5-6}$ als Modellsystem für Untersuchungen der C-Domäne herangezogen. Dessen zuvor festgestellte unerwartete Fähigkeit, das an PCP_{TycC5} gebundene Hexapeptid DPhe-Pro-Phe-DPhe-Asn-Gln zu zyklisieren, wurde eingehender studiert. Mit Hilfe der MS^n-Spektrometrie konnte die Kopf-zu-Schwanz-artige Konnektivität des zyklischen Produkts nachgewiesen werden. Die Substratspezifität dieser Reaktion wurde darüber hinaus anhand sechs weiterer Oligopeptide untersucht, von denen drei vom Enzym umgesetzt wurden. Desweiteren wurden Mutationsstudien durchgeführt, um die Gültigkeit bestehender Katalysemodelle für die C-Domäne zu prüfen. Nach einem dieser Modelle spielt ein konservierter Histidinrest eine entscheidende Rolle als Base bei der C-Domänenkatalyse. Obgleich die produzierten Alanin- und Valinmutanten dieses Aminosäurerestes zu korrekt gefalteten Proteinen führten, war doch deren Fähigkeit, Oligopeptide zu zyklisieren, verloren gegangen.

Durch Kristallisationsexperimente mit dem Wildtyp-Enzym *apo*-PCP-$C_{TycC5-6}$ konnte eine viel versprechende Bedingung gefunden werden, welche in Kollaboration mit der auf Kristallographie spezialisierten Arbeitsgruppe von Prof. Dr. Essen (Marburg) optimiert wurde. Es gelang, die erste Kristallstrukur eines bidomänalen nichtribosomalen Enzyms aufzuklären, wodurch bislang unbekannte Aspekte der Kommunikation zwischen einzelnen Domänen beleuchtet werden konnten. Die relative Anordnung der Domänen zueinander wurde als ein Zustand interpretiert, in dem *apo*-PCP eine Interaktion mit entweder anderen NRPS-Domänen oder einer Phosphopantethein-Transferase eingeht. Die hochvariable *linker*-Region, die beide Einheiten miteinander kovalent verknüpft, wurde erwartungsgemäß ohne Sekundärstruktur vorgefunden, obgleich schwache Wechselwirkungen mit den Oberflächen der beiden Domänen existieren. Kurioserweise findet sich ein vom verwendeten Puffer stammendes Sulfat-Ion in direkter Nähe zum vermuteten aktiven Zentrum der C-Domäne. In dieser Arbeit wird die Hypothese aufgestellt, dass dieses tetraedrische Anion im Kristall deswegen fixiert ist, weil es dem Übergangszustand während der Kondensationsreaktion ähnelt. Als Konsequenz hieraus ergibt sich, dass die C-Domäne den Übergangszustand durch ihre elektrostatische Umgebung stabilisiert, anstatt die Reaktion durch chemische Basenkatalyse voranzutreiben.

Table of Contents

1. Introduction 13

1.1 The Logic of Nonribosomal Peptide Assembly 13

1.1.1 The Essential Set of Domains 16

1.1.1.1 The Adenylation Domain 16

1.1.1.2 The Peptidyl Carrier Protein Domain 17

1.1.1.3 The Condensation Domain 18

1.1.2 Modifying Domains 20

1.2 Linear Gramicidin 23

1.3 A Closer Look at the Condensation Domain's Catalysis 25

1.4 Objectives 28

2. Abbreviations 29

3. Materials 32

3.1 Chemicals, Enzymes, and General Materials 32

3.2 Equipment 33

3.3 Vector Systems 34

3.3.1 pQE60 and pQE61 vectors 34

3.3.2 pBAD202/D-TOPO vector 35

3.3.3 pREP4 helper plasmid 36

3.4 Microorganisms 36

3.4.1 *E. coli* XL1-Blue 36

3.4.2 *E. coli* TOP-10 36

3.4.3 *E. coli* BL21 (DE 3) 37

3.4.4 *E. coli* BL21 (M15) [pREP4] 37

3.5 Media 37

3.6 Buffers 38

4. Methods 39

4.1 Construction of Recombinant Plasmids 39

4.1.1 pBAD202-TOPO[F-A-PCP_{LgrA1}]

and pBAD202-TOPO[F-A-PCP-C-A-$PCP_{LgrA1-2}$] 39

4.1.2 pQE61[PCP_{TycC5}], pQE61[PCP_{TycC5}] and pQE61[C_{TycC6}] 39

4.1.3 pQE61[PCP-$C_{TycC5-6}$] H224A and H224V mutants 40
4.2 Construction of Expression Strains 40
4.3 Protein Expression 41
4.4 Protein Purification 41
4.5 Synthesis of Aminoacyl- and Peptidyl-CoA Substrates 42
4.6 Synthesis of Transition State Analogs 43
4.7 Synthesis of N^{10}-Formyl-Tetrahydrofolate 49
4.8 ATP/PP_i-Exchange Assays 49
4.9 F Domain Assays 50
4.10 C Domain Assays 52
4.10.1 Cyclization Assays with PCP-$C_{TycC5-6}$ and the H224A and H224V Mutants 52
4.10.2 *In trans* Experiments with PCP_{TycC5} and C_{TycC6} 52
4.10.3 Diketopiperazine Formation Assay 52
4.11 Analysis with LCMS 53
4.12 Crystallization of *apo*-PCP-$C_{TycC5-6}$ 53
5. Results 54
5.1 The Formylation Domain 55
5.1.1 Selected Constructs 56
5.1.2 Cloning, Expression, and Purification 56
5.1.3 N^{10}-Formyl-Tetrahydrofolate Synthesis 57
5.1.4 The Specificity of the Adenylation Domain A_{LgrA1} 59
5.1.5 Formyltransferase Assays using F-A-PCP_{LgrA1} 60
5.1.6 Formyl Acceptor Substrate Specificity 61
5.1.7 Formyl Donor Substrate Specificity 62
5.1.8 The Necessity of a Formylated Starter Unit in Linear Gramicidin Biosynthesis 63
5.2 The Condensation Domain 65
5.2.1 Selected Constructs 65
5.2.2 Cloning, Expression, and Purification 66
5.2.3 Synthesis of Aminoacyl/Peptidyl-CoAs 67
5.2.4 Elongation/Cyclization Assays with PCP-$C_{TycC5-6}$ 67
5.2.5 Core 3 Mutants in Cyclization Experiments 70

5.2.6 Crystallization of PCP-$C_{TycC5-6}$ 71
5.2.7 The Structure of the Bidomain Enzyme 72
5.2.7.1 The Condensation Domain's Substructure 73
5.2.7.2 The PCP Domain's Substructure 74
5.2.7.3 The Linker Region 75
5.2.8 Design and Synthesis of Potential Inhibitors for the Condensation Domain 76
5.2.9 Inhibitory Assays Using TycA and TycB1 79
5.2.10 Inhibitory Assays Using PCP_{TycC5} and C_{TycC6} *in trans* 80
6. Discussion 83
6.1 The F Domain 85
6.1.1 The F Domain in the World of Formyltransferases 85
6.1.2 Acceptor Substrate Specificity of the Nonribosomal F domain F_{LgrA} 89
6.1.3 The Necessity of Formylation in Linear Gramicidin 89
6.2 The C Domain 90
6.2.1 Comparison to the Ribosomal Peptidyl Transferase Site 95
6.2.2 The Transglutaminase Homolog AdmF 98
6.3 Inter-Domain Communication 100
7. References 105
8. Appendix 111
9. Acknowledgements 115

Inhaltsverzeichnis

1. Einleitung 13

1.1 Die Logik nichtribosomaler Peptidsynthese 13

1.1.1 Die essenziellen Domänen 16

1.1.1.1 Die Adenylierungsdomäne 16

1.1.1.2 Die *Peptidyl-Carrier-Protein*-Domäne 17

1.1.1.3 Die Kondensationsdomäne 18

1.1.2 Modifizierende Domänen 20

1.2 Lineares Gramicidin 23

1.3 Die Katalyse der Kondensationsdomäne 25

1.4 Aufgabenstellung 28

2. Abkürzungen 29

3. Material 32

3.1 Chemikalien, Enzyme und Materialien 32

3.2 Ausstattung 33

3.3 Vektorsysteme 34

3.3.1 pQE60- und pQE61-Vektoren 34

3.3.2 pBAD202/D-TOPO-Vektor 35

3.3.3 pREP4-Helferplasmid 36

3.4 Mikroorganismen 36

3.4.1 *E. coli* XL1-Blue 36

3.4.2 *E. coli* TOP-10 36

3.4.3 *E. coli* BL21 (DE 3) 37

3.4.4 *E. coli* BL21 (M15) [pREP4] 37

3.5 Medien 37

3.6 Puffersysteme 38

4. Methoden 39

4.1 Konstruktion rekombinanter Plasmide 39

4.1.1 pBAD202-TOPO[F-A-PCP_{LgrA1}]

und pBAD202-TOPO[F-A-PCP-C-A-$PCP_{LgrA1-2}$] 39

4.1.2 pQE61[PCP_{TycC5}], pQE61[PCP_{TycC5}] und pQE61[C_{TycC6}] 39

4.1.3 pQE61[PCP-$C_{TycC5-6}$] H224A und H224V Mutanten 40
4.2 Konstruktion von Expressionsstämmen 40
4.3 Proteinexpression 41
4.4 Proteinreinigung 41
4.5 Synthese von Aminoacyl- und Peptidyl-CoA-Substraten 42
4.6 Synthese von Übergangszustandanaloga 43
4.7 Synthese von N^{10}-Formyl-Tetrahydrofolate 49
4.8 ATP/PP_i-Austausch-Assays 49
4.9 F-Domänen-Assays 50
4.10 C-Domänen-Assays 52
4.10.1 Zyklisierungsassays mit PCP-$C_{TycC5-6}$ und den H224A und H224V Mutanten 52
4.10.2 *In-trans*-Experimente mit PCP_{TycC5} und C_{TycC6} 52
4.10.3 Diketopiperazin-Bildungs-Assay 52
4.11 LCMS-Analyse 53
4.12 Kristallisation von *apo*-PCP-$C_{TycC5-6}$ 53
5. Ergebnisse 54
5.1 Die Formylierungsdomäne 55
5.1.1 Ausgewählte Konstrukte 56
5.1.2 Klonierung, Expression und Reinigung 56
5.1.3 Synthese von N^{10}-Formyl-Tetrahydrofolat 57
5.1.4 Die Spezifität der Adenylierungsdomäne A_{LgrA1} 59
5.1.5 Formyltransferase-Assays mit F-A-PCP_{LgrA1} 60
5.1.6 Formylakzeptor-Substratspezifität 61
5.1.7 Formyldonor-Substratespezifität 62
5.1.8 Die Notwendigkeit einer formylierten Starteinheit für die Biosynthese des linearen Gramicidins 63
5.2 Die Kondensationsdomäne 65
5.2.1 Ausgewählte Konstrukte 65
5.2.2 Klonierung, Expression und Reinigung 66
5.2.3 Synthese von Aminoacyl/Peptidyl-CoA-Substraten 67
5.2.4 Elongations-/Zyklisierungs-Assays mit PCP-$C_{TycC5-6}$ 67
5.2.5 Core-3-Mutanten in Zyklisierungsexperimenten 70

5.2.6 Kristallisation von PCP-$C_{TycC5-6}$ 71
5.2.7 Die Struktur des Bidomänenenzyms 72
5.2.7.1 Die Substruktur der Kondensationsdomäne 73
5.2.7.2 Die Substruktur der PCP-Domäne 74
5.2.7.3 Die *linker*-Region 75
5.2.8 Entwurf und Synthese potenzieller Inhibitoren für die Kondensationsdomäne 76
5.2.9 Inhibitorische Assays mit TycA und TycB1 79
5.2.10 Inhibitorische Assays mit PCP_{TycC5} und C_{TycC6} *in trans* 80
6. Diskussion 83
6.1 Die F-Domäne 85
6.1.1 Die F-Domäne im Vergleich zu anderen Formyltransferasen 85
6.1.2 Akzeptorsubstrat-Spezifität der nichtribosomalen F-Domäne F_{LgrA} 89
6.1.3 Die Notwendigkeit der Formylierung für lineares Gramicidin 89
6.2 Die C-Domäne 90
6.2.1 Vergleich mit der ribosomalen Peptidyltransferase 95
6.2.2 Das Transglutaminase-Homolog AdmF 98
6.3 Kommunikation zwischen den Domänen 100
7. Literaturverzeichnis 105
8. Anhang 111
9. Danksagung 115

1. Introduction

1.1 The Logic of Nonribosomal Peptide Assembly

Nonribosomal peptide synthetases (NRPSs) comprise a unique class of multi-domain enzymes capable of producing peptides [Walsh 2004, Finking 2004, Fischbach 2006, Keller 2003]. They are found among a broad variety of fungi, bacteria, and lower eucaryotes. In contrast to the ribosomal machinery, where the mRNA template is translated to proteins, the order of catalytically active entities within these synthetases intrinsically determines the sequence of building blocks in the peptide product (Figure 1.1).

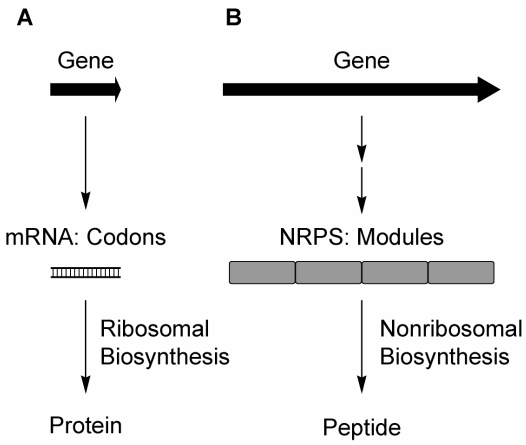


Figure 1.1: Comparison of ribosomal and nonribosomal peptide synthesis. (**A**) In the ribosomal information pathway, the sequence of codons in the mRNA determines the sequence of amino acids in the protein whereas (**B**) the sequence of modules in the nonribosomal peptide synthetases determines the primary sequence of the peptide product.

As a consequence, generally speaking, each NRPS can only produce one defined peptide product. This is chemically implemented by the fact that all substrates and reaction intermediates are spatially fixed to the synthetase by covalent linkage – thereby eliminating side product formation due to diffusion. The catalytic entities responsible for the incorporation of a distinct building block into the product are called modules. Each module carries out several chemical steps required for the synthesis of nonribosomal peptides: recognition of the building block, activation, covalent attachment, translocation, and condensation. In several cases, additional modifications are found, such as epimerization (tyrocidine, [Stein 2006]), cyclization (gramicidin S, [Kohli 2001]), oxidation (epothilone, [Chen 2001]), reduction and formylation (linear gramicidin, [Schracke 2005]), and methylation (cyclosporin, [Velkov 2003]). *In vitro* studies have shown that each module

can further be subdivided into catalytically active domains to which the different reactions mentioned above can be assigned [Walsh 2004]. Thus, the so-called adenylation (A), peptidyl carrier protein (PCP), and the condensation (C) domains were identified as being essential to all NRPSs. Besides, a second group of so-called optional domains exists: the epimerization (E), cyclization (TE or Cy), oxidation (Ox), reduction (R), N-methylation (Mt) and formylation (F) domains. Aside from NRPSs themselves, several enzymes are known to act on some of the peptides while they are still bound to the synthetase or even after their release. These modifying enzymes can glycosylate, halogenate (both Vancomycin, [Hubbard 2003]) or reduce (linear gramicidine, [Schracke 2005]) the peptides *in trans*. With several hundred different building blocks found in nonribosomal peptides, it becomes evident that their diversity is just as vast (Figure 1.2) as their biological functions – ranging from cytostatic, immunosuppressive, antibacterial to antitumor properties.

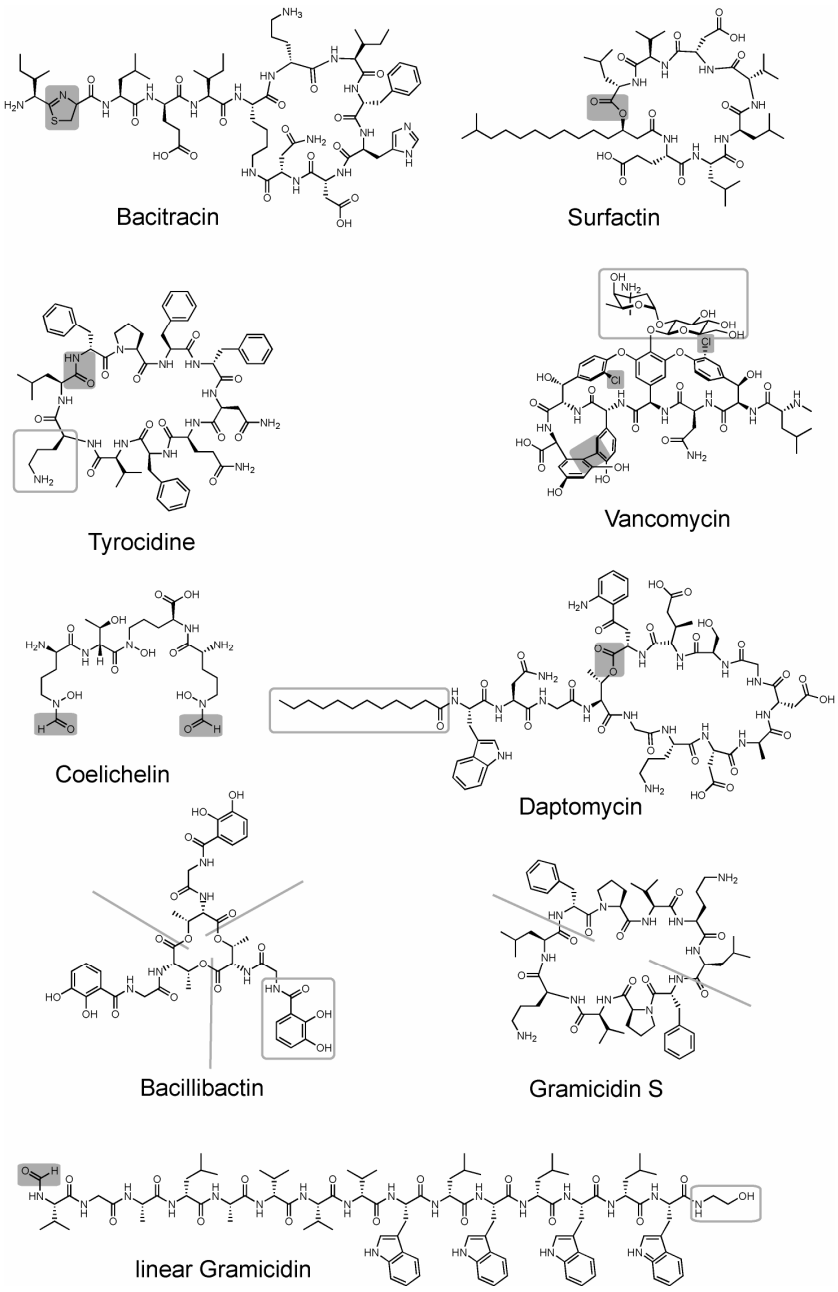


Figure 1.2: Examples for the diversity of nonribosomal peptides. A selection of unusual building blocks, chemical and structural features that contribute to the vast diversity of this class of metabolites are highlighted: Heterocycle (bacitracin); lactone (surfactin, daptomycin), ornithine and lactam (tyrocidine); sugar, chlorinated aromats, C-C crosslink (vancomycin); N-formyl groups (coelichelin and linear gramicidin); fatty acid (daptomycin); dihydroxybenzoate and trimeric organization (bacillibactin); dimeric organization (gramicidin S); ethanolamine (linear gramicidin).

The catalytically active entities that NRPSs are composed of can be classified as being either essential to all NRPSs or responsible for special modifications. Only when a set of domains correctly acts in appropriate order, the designated product can be synthesized [Fischbach 2006]. The mechanistic functions and interactions of these domains are described in the following sections.

1.1.1 The Essential Set of Domains

Three nonribosomal domains are considered to be essential: the A, PCP, and C domains. The knowledge of the reactions catalyzed by these domains is a prerequisite to understanding the mechanics of nonribosomal peptide assembly.

1.1.1.1 The Adenylation Domain

Before any peptide formation can occur, the amino acids, or, generally speaking, the building blocks to be condensed need to be recognized and activated [Luo 2001]. The adenylation (A) domains (~550 aa) are capable of specifically binding one such building block. Once bound, the same enzyme catalyzes the formation of the corresponding acyl-adenylate-monophosphate by consumption of ATP. The resulting mixed anhydride is the reactive species that can further be processed by the NRPS machinery (figure 1.3). Sequence alignments, mutational studies, and structural data have revealed that amino acid residues at certain positions in the enzyme determine the specificity of an A domain [Stachelhaus 1999]. This can be explained by the thereby generated chemical and physical environment of the substrate binding site. Some A domains, however, are known to have a relaxed substrate specificity. In these cases, chemically or sterically similar amino acids are also recognized, analogously processed and thus found at that very position in the product. For example, the A domain of the first module of the gramicidin synthetase LgrA [Schoenafinger 2005] not only activates valine but also isoleucine, which is found in 5 % of linear gramicidins extracted from producing strains.

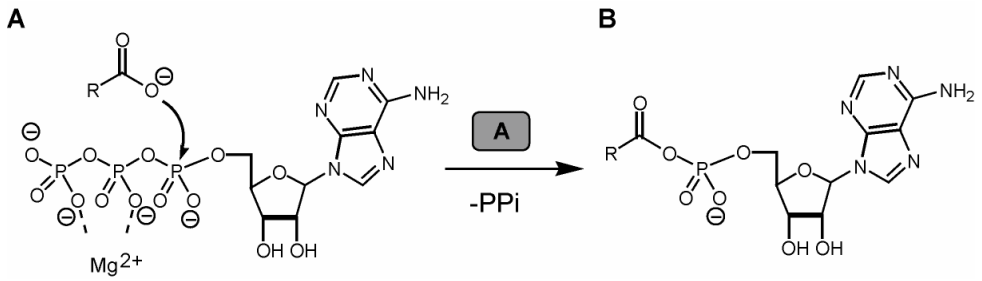


Figure 1.3: The A domain recognizes building blocks for the nonribosomal peptide synthesis and activates their carboxyl groups **(A)** by generating a mixed anhydride **(B)** upon consumption of ATP.

1.1.1.2 The Peptidyl Carrier Protein Domain

When the first building block has been recognized and activated by the A domain, the next essential domain comes into play: the peptidyl carrier protein (PCP) domain. Like the acyl carrier protein (ACP) in fatty acid biosynthesis, this PCP domain is responsible for keeping the reaction intermediates bound to the enzymatic machinery. Thus, a directed order of further reaction steps can be implemented by controlled translocation, and NRPSs are thus often described as assembly line-like machineries. The PCP domain consists of approximately 90 amino acid residues and is believed to re-arrange itself to at least three different tertiary structures, as necessary for interaction with the surrounding domains at certain stages of synthesis [Koglin 2006]. Just like ACPs, the PCP domains are also dependent on a post-translational modification to function. This modification is the attachment of a 4'-phosphopantetheine (Ppan) cofactor to a conserved serine residue (figure 1.4). The reaction is catalyzed by a Ppan transferase, such as Sfp from the surfactin NRPS system. It is a prerequisite for all functional NRPSs, and depending on whether this modification has been made, NRPSs are therefore classified as being in their *apo*- or *holo*-state.

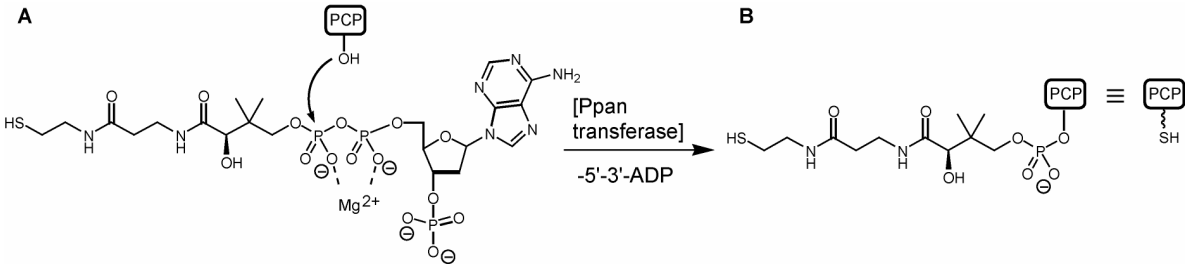


Figure 1.4: The *apo*-to-*holo* conversion of a PCP domain is catalyzed by Ppan transferases. **(A)** The *apo*-PCP domain's active site serine residue attacks the β-phosphate group of CoA, leading to *holo*-PCP **(B)**. A common schematic abbreviation for *holo*-PCP is shown, which is used throughout this work.

The terminal thiol group of this cofactor is the nucleophile which attacks the mixed anhydride (acyl-AMP) generated by A domains and covalently binds the NRPS substrates *via* a thioester bond (figure 1.5). After such an acylation, the PCP domain directs the substrate towards the next processing domain. If we leave out any optional modifying domains at this point, this next domain would generally be a condensation (C) domain.

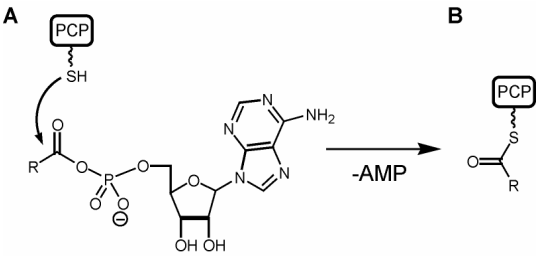


Figure 1.5: Nonribosomal aminoacylation of *holo*-PCP. **(A)** The Ppan cofactor's terminal thiol group attacks the activated carboxyl group of a building block activated by the A domain. **(B)** By elimination of 5'-AMP the acyl-*holo*-PCP is formed.

1.1.1.3 The Condensation Domain

The C domain (~450 aa) is needed for the condensation of two biosynthetic intermediates during nonribosomal peptide assembly [Belshaw 1999, Keating 2002]. The PCP-bound electrophilic donor substrate is presented from the N-terminal side of the synthetase. On the other side, the nucleophilic acceptor substrate – analogously bound to the PCP domain of the next module – reaches back to the active site of the C domain from the other direction (downstream). In the first condensation reaction of an NRPS, both these substrates would typically be aminoacyl groups connected to their PCP domains (figure 1.6).

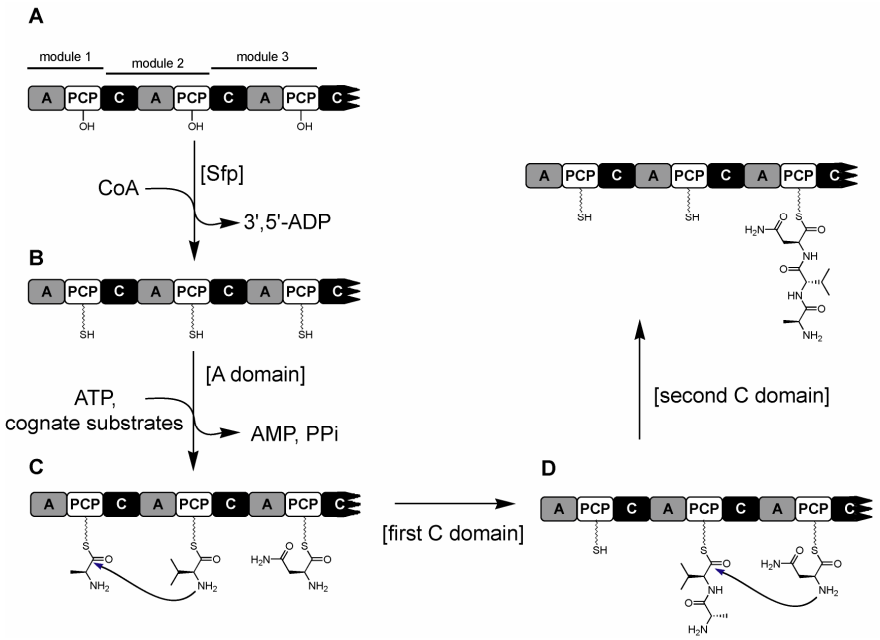


Figure 1.6: Single reactions in NRPSs and their timing. **(A)** After ribosomal synthesis of the *apo*-enzymes, the PCP domains are post-synthetically modified with 4'-Phosphopantetheine cofactors by a 4'Ppan transferase, e. g. Sfp. **(B)** In a second step, the A domains bind their cognate substrates as well as ATP and form the corresponding acyl-adenylate intermediates. These are transferred onto the cofactor of the neighboring PCP domains. **(C)** The C domains catalyze the condensation of two building blocks. The specificities of C domains and the affinities of aminoacyl-/peptidyl-*holo*-PCP domains ensure that no internal start reactions occur: (**D**) Only after the first condensation domain has acted, the second C domain further processes the intermediate. During synthesis, the growing product chain is continuously translocated towards the C-terminal end of the enzyme.

Condensation is initiated by the nucleophilic attack of the α-amino group of the acceptor substrate onto the thioester group of the donor substrate (detailed information is given in chapter 1.3). The cofactor of the upstream PCP domain is released, and the resulting amide bond now belongs to the dipeptide which remains bound to the downstream PCP domain. Thus, a translocation of the condensation product towards the next module has occurred. All condensation reactions are strictly unidirectional – always transporting the growing product chain towards the module closer to the C-terminus of the machinery. The elongated peptide then serves as donor in a subsequent condensation step on the next module. The intrinsic structural interactions and affinities that direct the synthesis are not yet understood.

Usually, there are as many condensation domains in an NRPS as there are peptide bonds in the linear peptide product. This general translocation model implies that the biosynthesis is linear – altogether dependent on delicate, situationally changing affinities that guarantee correct timing for each single reaction and that prevent side reactions. Even though this model successfully puts the biosynthetic enzymes in relation with their products for the majority of known NRPS systems, some exceptions are known: The structures of syringomycin [Bender 1999] or coelichelin [Lautru 2005] (Figure 1.2) cannot be sufficiently explained by merely deciphering the build-up of their NRPSs when using this model. Obviously, there are other regulatory mechanisms and forms of inter-domain communication which remain to be elucidated.

When the last condensation reaction has occurred, the linear precursor needs to be released from the enzyme. For this important last step, several mechanisms are known: simple hydrolysis of the thioester (balhimycin, vancomycin), intramolecular cyclization leading to a lactam (tyrocidine, bacitracin) or a lactone (surfactin), or even reductive thioester cleavage (linear gramicidin). In some cases, the linear precursor is dimerized (gramicidin S) or even trimerized (bacillibactin, enterobactin) prior to cyclization (Figure 1.2). Even though these reactions are critical for the compounds' bioactivities, the catalytic domains responsible for the release are not found in all NRPS systems and are therefore called "modifying" domains.

1.1.2 Modifying Domains

Apart from the essential domains in NRPSs, there are a number of so-called modifying domains which are not found in every NRPS system. Nevertheless, they are required for proper processing of their designated substrate within their synthetase. Deletion or inactivation of these modifying domains usually results in the production of compounds with bioactivities severely reduced or altogether abolished.

The majority of nonribosomal peptides have a cyclic connectivity. In these cases, a C-terminal so-called thioesterase (TE) domain is often found in the synthetase. These TE domains all share an invariant serine residue belonging to a catalytic triad (Asp-His-Ser) which is known to be acylated with the linear peptide prior to cyclization [Kohli 2001]. Once the substrate is thus translocated from the PCP domain onto the TE domain, the regio- and stereospecific intramolecular attack of a nucleophile onto the C-terminal carbonyl group of the substrate is directed by the enzyme. This nucleophile can be the N-terminal α-amino group of the linear peptide (tyrocidine, gramicidin S), a side chain amino (bacitracin) or hydroxyl group (surfactin). Since the ester bond between the substrate and the TE domain is cleaved by these cyclization reactions, the resulting lactams or lactones are released from the synthetic machinery by this step. In a few cases, the modular arrangement of NRPSs suggests that only half (gramicidin S) or a third (bacillibactin, enterobactin) of the extracted peptide product can be produced by one assembly line-like synthesis. These synthetases are considered iterative [Gehring 1998] since they have to complete more than one linear peptide synthesis before one molecule of the secondary metabolite can be released. According to a proposed model, the first precursor is translocated onto the TE domain, the second monomer is then produced and transferred to the TE domain-bound first monomer leading to a dimer. An analogous trimerization occurs – if applicable – and finally the product is released by cyclization.

Another modifying reaction which is commonly found in NRPSs is the epimerization of an amino acid [Stein 2006]. Epimerization (E) domains, which are always situated directly downstream of a PCP domain, catalyze these reactions. The most C-terminal amino acid of the reaction intermediate is racemized by an E domain, no matter if the substrate is an aminoacyl group alone or a peptidyl group. The mechanism of these epimerization

domains is so far unclear even though a catalysis which involves one or more catalytic bases to deprotonate the α-carbon atom as a first step seems likely. The resulting planar double bond species then needs to be reprotonated from the other side to invert the absolute configuration of the building block. This can be accomplished by a nearby protonated catalytic base in the enzyme or water which is positioned opposite of the first catalytic base. Nevertheless, always a mixture of both stereoisomers can be detected when the substrate bound to the enzyme is analyzed – indicative for either a non-stereoselective or a reversible reaction. Once the epimerized substrate undergoes the subsequent condensation reaction, only the species with inverted stereocenter is found in the elongated product. Thus, the C domain only processes the inverted species. In some rare cases, the C domain itself also exhibits epimerization activity besides its normal function, and it is then called "dual C/E" domain (arthrofactin, [Balibar 2005]).

Another structural feature often found in NRPS products is N-methylated amide bonds. The domain that introduces this C_1 unit, the so-called N-methyltransferase (Mt) domain, is situated between the A and PCP domain [Billich 1987]. By consumption of S-adenosyl-methionine (SAM) the α-amino group of the acceptor substrate is methylated prior to condensation with the donor.

In the case of linear gramicidin, the N-terminus of the nonribosomal peptide carries a formyl group (see chapter 1.2). Thus, one can find a distinct putative formylation (F) domain at the very N-terminus of the synthetase. Another formylated NRPS product is coelichelin whose ornithine residues are believed to be N_δ-formylated *in trans* by a formyltransferase genetically associated with the NRPS [Lautru 2005].

The essential condensation domain mentioned above can – in some cases – not only condensate but also catalyze a side chain cyclization. It is then called cyclization (Cy) domain. The cyclization is initiated by a nucleophilic attack of the side chain heteroatom on the carbonyl group of the amide bond formed by the same domain. When water is eliminated, a stable pentacycle has been integrated into the peptide chain without altering the rest of the backbone. The nucleophiles known to be reactants in these Cy domain reactions are either threonine/serine (mycobactin A, [Quadri 1998]) or cysteine (bacitracin,

[Eppelmann 2001]). The former leads to (methyl-)oxazoline heterocycles while the latter gives rise to thiazoline-like units. Another domain sometimes associated with this heterocyclization is the oxidation (Ox) domain (epothilone, [Chen 2001]). It is located between A and PCP domains and it catalyzes the oxidation of oxa/thiazoline intermediates, leading to oxazoles or thiazoles, respectively.

1.2 Linear Gramicidin

Linear gramicidin is a linear nonribosomal secondary metabolite produced by *Bacillus brevis* ATCC 8185 during sporulation phase [Hotchkiss 1940]. Its primary structure can be described as sequence of 15 hydrophobic amino acids, which exhibit a striking alternation of D- and L-configured side chains (figure 1.7). Moreover, both termini carry unusual modifications: The first amino acid is found N_{α}-formylated, and the last amino acid is connected to an ethanolamine unit *via* an amide bond. When extracted from cultures, a characteristic mixture of different isoforms is found [Sarges 1965], as depicted in the following figure 1.7:

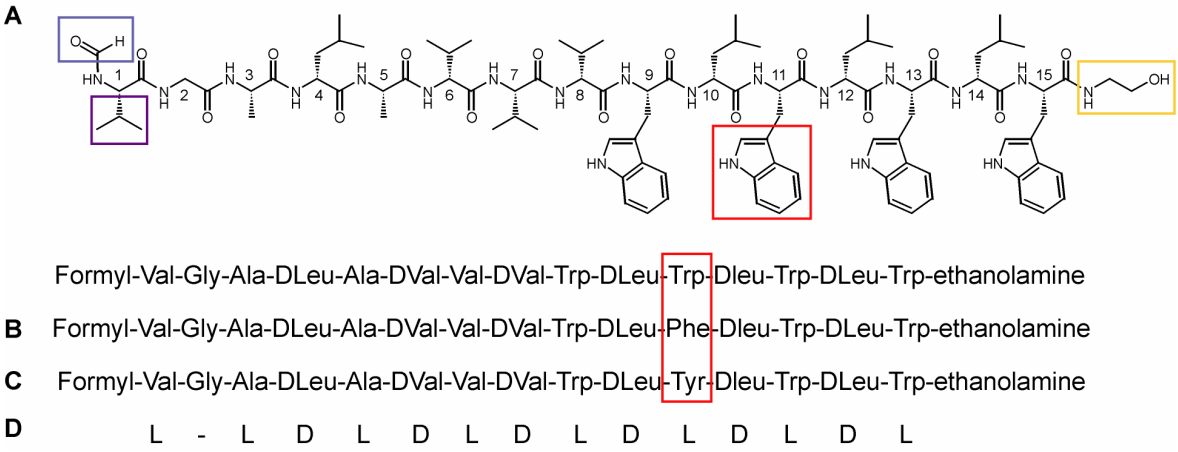


Figure 1.7: The primary structure of linear gramicidin and its known isoforms. **(A)** Structure formula of linear gramicidin A. The N-terminal formyl group (blue) and the C-terminal ethanolamine (yellow) are highlighted. The valine residue at position 1 (purple) is found to be substituted by isoleucine in 5 % of all gramicidin isoforms. **(B)** In gramicidin B, Phe is found at position 11 (red) instead of Trp, **(C)** and the same residue is Tyr in gramicidin C. **(D)** Illustration of the alternating D- and L-configurations in the sequences.

The natural mixture of linear gramicidins is called gramicidin D. It typically consists of 80 % A, 5 % B, and 15 % of the isoform C. Furthermore, all of these variants have been reported to carry isoleucine residue at position 1 instead of valine in 5 % of their population [Sarges 1965].

The NMR structure of linear gramicidin A_{Val1} in presence of dodecyl phosphocholine micelles (figure 1.8) [Townsley 2001] shows its dimeric organization and right-handed β-(6.3)-helical secondary structure. The head-to-head dimer of linear gramicidin acts as an ion channel (3-4 Å wide) which is believed to serve its producer as a weapon against competitors. The hydrophobic side chains are pointing away from the channel axis while

the formyl groups at the dimer interface are part of a hydrogen network stabilizing the binary complex.

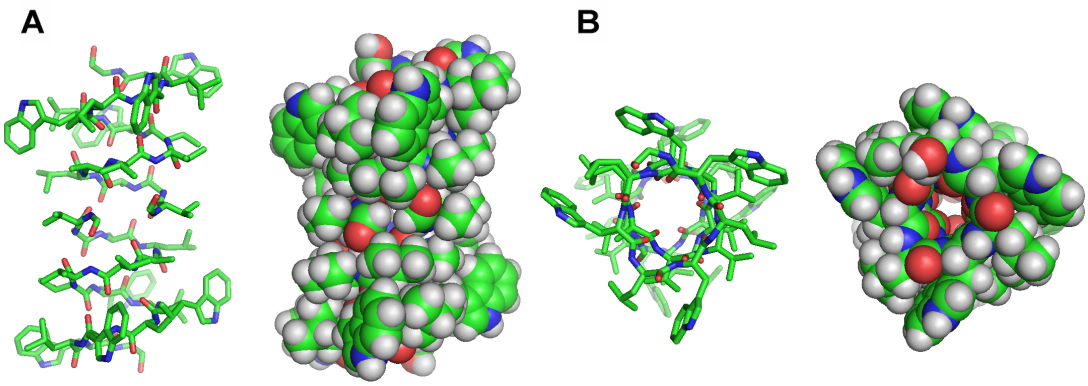


Figure 1.8: The NMR structure of linear gramicidin A_{Val1}. **(A)** side view: The formyl groups (arrows) are located at the interface of the dimeric complex. **(B)** The top view reveals the polar channel formed by the helical dimer.

The biosynthetic cluster responsible for the production of linear gramicidin, *lgr*, consists of five synthetases: LgrA-E (figure 1.9) [Kessler 2004]. With the characterization of the two terminal reductase (R) domains (LgrD4 and LgrE), the origin of ethanolamine at the C-terminus could be explained by reduction of the glycine residue incorporated by module 16 (LgrD4) [Schracke 2005]. The thioester group of the template-bound linear 16-mer is first reduced to the aldehyde by R_{LgrD4}, and subsequently reduced once more by LgrE *in trans* – giving rise to the alcohol.

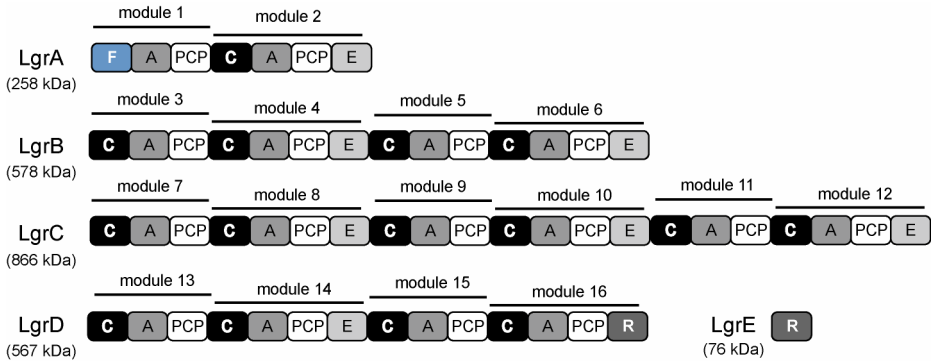


Figure 1.9: The *lgr* cluster encodes for five synthetases, consisting of 16 modules with 56 catalytic domains in total. The F domain (blue) is located at the N-terminus of LgrA.

The N-terminal domain of LgrA (~200 aa), however, was hypothesized to be a formyltransferase (F) domain incorporating the formyl group at position 1 – an assumption that was supported by sequence comparison with other formylating enzymes, such as

bacterial methionyl-tRNA-formyltransferases (33-35% identity, and 52-58% similarity) [Schracke 2005a]. In this work, the biochemical characterization of the F domain is presented.

1.3 A Closer Look at the Condensation Domain's Catalysis

Special attention is drawn to the C domain here, as it is subject to substantial parts of this work.

The C domain generally catalyzes the condensation reaction between two PCP-bound substrates. The donor substrate's C-terminus is activated as a thioester which also serves as the covalent tether to the enzymatic template prior to the reaction. The acceptor substrate's free (α-)amino group is the nucleophile which attacks the donor's C-terminus. As a result, a tetrahedral intermediate in accordance to the addition/elimination reactions of oxoester chemistry is generated, which eventually leads to the thermodynamically more stable amide/peptide bond in the elongated product. The necessity of a C domain to catalyze such a reaction under physiological conditions has been pointed out before [Stachelhaus 1998, Bergendahl 2002]. However, the mode of its action is not yet fully understood.

Characteristic sequences in the primary structure, so-called core motifs, have been determined for C domains, and the core 3 motif "HHxxxDG" was suggested to be involved in the catalytic process. Upon mutation of the second histidine (H147) in this motif to valine, the condensational activity of the C domain of TycB1 (tyrocidine NRPS, *Bacillus brevis* ATCC 8185) was abolished [Stachelhaus 1998]. Consequently, a mechanistic model was proposed in which this histidine was suggested to act as a catalytic base designated to capture the excess proton of the acceptor group after the attack on the acceptor carboxyl [Bergendahl 2002] (figure 1.10). Nevertheless, results contrary to this were obtained when the stand-alone C domain VibH from the vibriobactin nonribosomal system of *Vibrio cholerae* was investigated [Keating 2002]. The mutation of the analogous histidine residue to alanine led to only 90 % loss of activity, indicating that the histidine residue is not essential in a mechanistic sense.

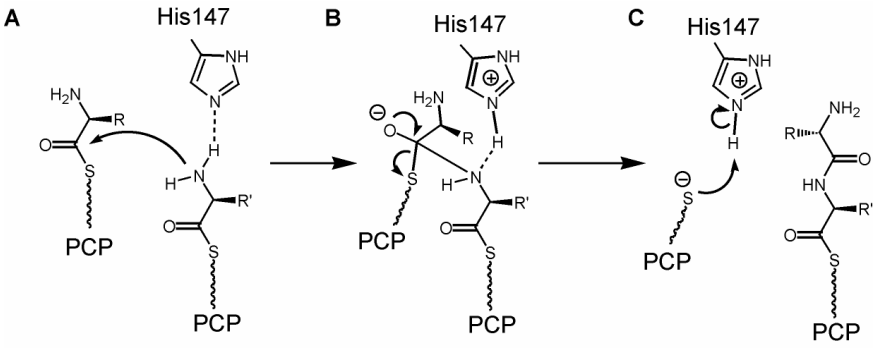


Figure 1.10: The proposed role of histidine 147 as a base catalyst during peptide bond formation with C_{TycB1}. **(A)** the nucleophilic attack is facilitated by partial abstraction of a proton. **(B)** The elimination of the donor substrate's thiolate group leads to the elongated product **(C)**, and the proton is transferred back to the Ppan cofactor.

Instead, it was suggested, that the positioning of the reactants alone is sufficient for turnover. However, the reaction catalyzed by VibH is somewhat atypical, as it condenses PCP-bound dihydroxybenzoate and free norspermidine. The first C domain structure – presented in the same publication [Keating 2002] – shows a pseudo-dimeric organization of two chloramphenicol-transferase monomer homolog subdomains, which gives VibH a V-shaped overall structure (figure 1.11). The conserved histidine residue mentioned above is situated in the middle of the large cleft defined by the V-shape which is believed to be the substrate channel.

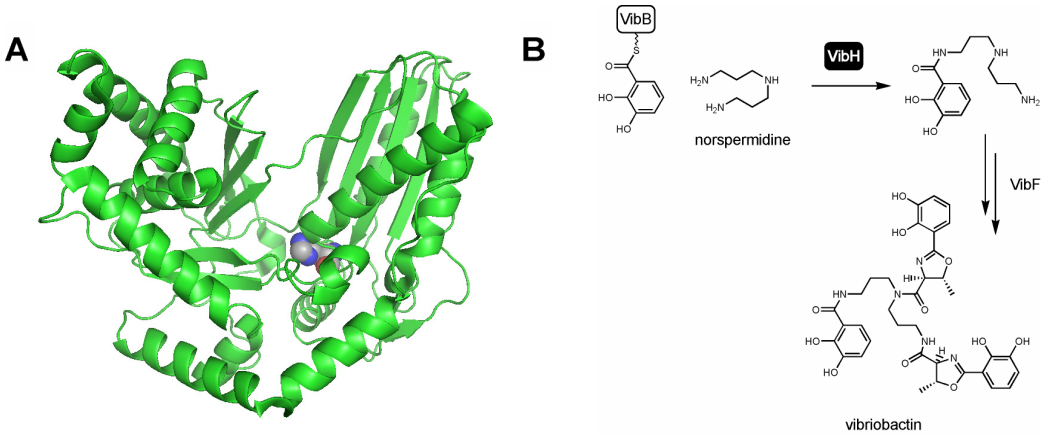


Figure 1.11: (A) The structure of VibH (PDB accession number 1L5A). Two subdomains (left and right) are arranged, so that a V-shape overall structure with a cleft in its middle is formed. The conserved second histidine residue of the C domains' core motif 3 is show as sphere model. **(B)** The reaction catalyzed by the stand-alone C domain VibH. Dihydroxybenzoate bound to the aryl carrier protein VibB is condensed with norspermidine to give rise to a biosynthetic intermediate during vibriobactin synthesis.

In 2003, however, mutational studies with the dissected C domain of EntF from the enterobactin NRPS, again indicated a critical role of the second histidine in the core 3 motif, since *in vitro* activity was found abolished upon its mutation [Roche 2003]. It has been speculated since, whether catalysis originates from substrate postioning, the presence of a base, or residues stabilizing the transition state. This matter is furthermore discussed in this work here.

Besides their ability to condensate two reactants, the C domains can further be categorized by additional functions and roles within their nonribosomal machinery (figure 1.12). The most basic type is the so called internal C domain. It interacts with two PCP domains *in cis*. The second type is the N-terminal C domain that communicates with a preceding synthetase which provides a PCP-bound donor substrate. This communication has been linked to short peptide sequences with predicted α-helical structures, which lie at the termini of both synthetases [Hahn 2004]. They are therefore called communication mediating (COM) domains. Typically, the donor substrate carries a D-configured C-terminus – a feature incorporated by a preceding E domain – and the C domain is then specific for D-configured donor substrates [Clugston 2004]. Furthermore, C domains can be located at the very beginning of a synthetase (initatior C domains), usually interacting with an acyl carrier protein to incorporate an N-terminal lipid moiety into the metabolite. In the cyclosporin synthetase, there is a C-terminal C domain, which is hypothesized to be responsible for the macrolactamization of the linear full-length precursor.

As mentioned before, C domains can have unusual activities. In the bacitracin synthetase for example, the first C domain additionally exhibits the ability to cyclize its condensation product Ile-Cys *via* the cysteine side chain. As a result, the amide bond is converted into an imine, which is part of a thiazoline system. These domains are called cyclization (Cy) domains. Another unusual type of C domain possesses both epimerization and condensation activity. These “dual C/E” domains [Balibar 2005] show an additional histidine motif in the primary sequence (see appendix). A close relative of C domains are the E domains. Even though functionally different, they have a striking sequence similarity to C domains. Characteristic insertions are believed to be responsible for that (see appendix).

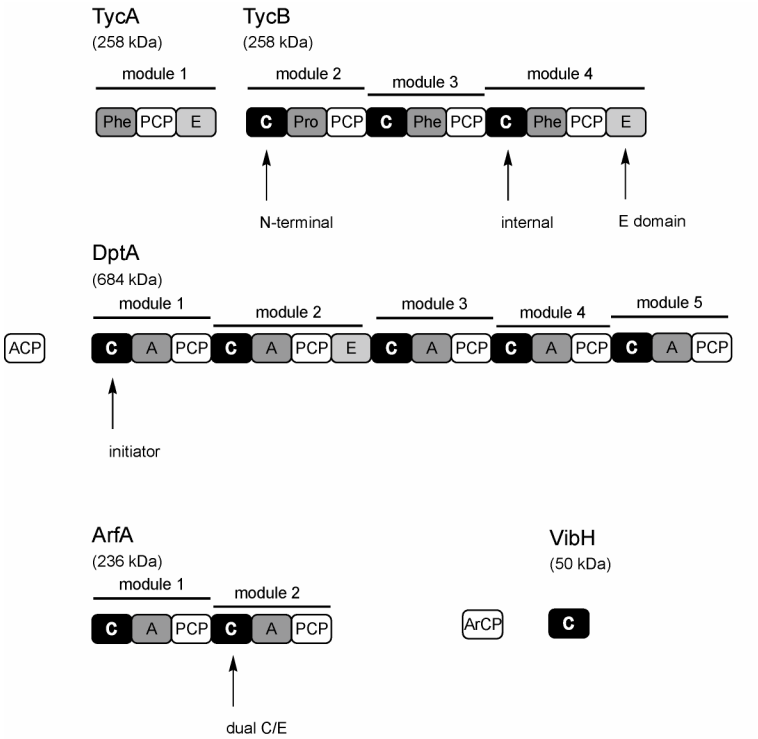


Figure 1.12: Examples of different C domain types. In TycB, an N-terminal and an internal C domain are indicated. Furthermore, the related C-terminal E domain is pointed out. In DptA (daptomycin NRPS), an initiator C domain interacts with an acyl carrier protein (ACP) for lipo-initiation. A dual C/E domain is indicated in the bimodular system ArfA which belongs to the arthrofactin NRPS system. The stand-alone C domain VibH only interacts with an aryl carrier protein (ArCP) which supplies the donor substrate.

1.4 Objectives

The main objective of this thesis was the investigation of catalytic processes performed by nonribosomal peptide synthetases that result in amide bonds. To this end, two model systems were chosen which perform different amide bond forming reactions.

Firstly, the N-terminal domain of LgrA with proposed N-formyltransferase activity was to be characterized, as this domain type had merely been hypothesized prior to this work. The biochemical characterization should provide answers to questions of substrate specificity and its mechanistic role within the synthetase.

Secondly, the condensation domain, ubiquitous in NRPSs, was subjected to further investigations. An unexpected lactamization reaction previously observed *in vitro* with a PCP-C bidomain enzyme from the tyrocidine NRPS should be further characterized. Moreover, deeper insights into the catalytic process of C domains were sought by a crystallographic approach and mutational studies.

2. Abbreviations

aa	amino acid
Ac	acetyl
AcOH	acetic acid
ACP	acyl carrier protein
A-domain	adenylation domain
Amp	ampicillin
AMP	adenosine-5'-monophosphate
ADP	adenosine-5'-diphosphate
ArCP	aryl carrier protein
ATP	adenosine-5'-triphosphate
B	base
Boc	tert-butyloxycarbonyl
bp	base pairs
BSA	bovine serum albumin
calc.	calculated
CAT	chloramphenicol/acetyl-transferase
C domain	condensation domain
CoA	coenzyme A
cP	cyclic peptide
CP	carrier protein
Cy domain	heterocyclization domain
Da	Dalton
DCC	dicyclohexylcarbodiimide
DCM	dichloromethane
DHB	dihydroxybenzoyl
DMSO	dimethyl sulfoxide
DIPEA	diisopropylethylamine
DMF	N,N-dimethylformamide
dNTP	2'-desoxynucleosid-5'-triphosphate
E domain	epimerization domain
EDTA	ethylene-diamino-tetraacetic acid
EK	enterokinase
Em	emission
ESI-MS	electron spray ionization – mass spectrometry
eq.	equivalent
F domain	formyation domain
FAS	fatty acid synthase
Fen	fengycin
Fig.	Figure
FMN	flavin mononucleotide
Fmoc	9-fluorenylmethyloxycarbonyl
FPLC	fast performance liquid chromatography
fTHF	formyl-tetrahydrofolate
HBTU	2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HEPES	2-N'-[N-(2-hydroxylethyl)-piperazinyl]-ethansulfonic acid
Hex	hexanoyl
HOBt	1-hydroxybenzotriazole
HPLC	high performance liquid chromatography

IPTG	isopropyl-β-D-thiogalactoside
Kan	kanamycin
kb	kilo base pairs
LB medium	Luria-Bertani/Lysogeny broth medium
LCMS	liquid chromatography-mass spectrometry
MALDI-TOF	matrix assisted laser desorption ionization-time of flight
MCS	multiple cloning site
MES	2-morpholinoethanesulfonic acid
MS	mass spectrometry
Mt domain	N-methyltransferase domain
n. d.	not detected
NMR	nuclear magnetic resonance
NRPS	nonribosomal peptide synthetase
NTA	nitrilotriacetate
OD	optical density
Ox domain	oxidation domain
PAGE	polyacrylamide gel electrophoresis
PCP	peptidyl carrier protein or thiolation domain
PCR	polymerase chain reaction
PKS	polyketide synthase
Ppan	4'-phosphopantetheine
PP_i	inorganic pyrophosphate
PyBOP	benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate
R domain	reductase domain
rpm	revolutions per minute
RT	room temperature
SAM	S-adenosylmethionine
S_B	streptogramin B
SDS	sodium dodecylsulfate
Sfp	4'-phosphopantetheine transferase involved in surfactin production
SNAc	N-acetylcysteamine
SPPS	solid phase peptide synthesis
Srf	surfactin
tBu	tert-butyl
TCEP	tris(carboxyethyl)phosphine
T domain	thiolation domain or peptidyl carrier protein
TE domain	thioesterase domain
TEOF	triethoxy-orthoformiate
TFA	trifluoroacetic acid
TFE	trifluoroethanol
THF	tetrahydrofolate
TIPS	triisopropylsilane
t_R	retention time
Tris	tris-(hydroxymethyl)-aminomethane
Trt	trityl
Tyc	tyrocidine synthetase
V	volts
v/v	volume per volume
wt	wild type
w/v	weight per volume

Amino acids:	**3-**	**1-letter code**
alanine	Ala	A
arginine	Arg	R
asparagine	Asn	N
aspartate	Asp	D
cysteine	Cys	C
glutamine	Gln	Q
glutamate	Glu	E
glycine	Gly	G
histidine	His	H
isoleucine	Ile	I
kynurenine	Kyn	U
leucine	Leu	L
lysine	Lys	K
methionine	Met	M
ornithine	Orn	O
phenylalanine	Phe	F
proline	Pro	P
serine	Ser	S
threonine	Thr	T
tryptophan	Trp	W
tyrosine	Tyr	Y
valine	Val	V

3. Materials

In this section, the chemical compounds, technical equipment, and other materials used in this work are listed.

3.1 Chemicals, Enzymes, and General Materials

Chemicals not stated here were purchased as standard compounds from other manufacturers in *p.a.*-quality.

Manufacturer	**Product**
Amersham Biosciences European GmbH	various restriction endonucleases, ampicillin, IPTG, kanamycin, yeast extract, coomassie brilliant blue G and R250, agar Nr.1, HiTrap™-desalting columns,
Applied Biosystems	ABI prism^{TM} dRhodamine terminator cycle sequencing ready reaction kit v. 3.0, HiDi Formamide
Bachem	N_α-Fmoc-protected amino acids, N_α-Boc-protected amino acids
Eurogentech	agarose, electroporation cuvettes
Fluka	SDS, TEMED, DMF
Macherey und Nagel	C_{18}-Nucleodur HPLC column, C_{18}-Nucleosil-NH_2 HPLC column
Merck	silica gel 60 F_{254} plates TLC
Millipore	dialysis membrane (0,025 µm)
New England Biolabs	desoxyribonucleotides (dATP, dTTP, dGTP, dCTP), prestained protein molmarker, various restriction endonucleases, 1kb-DNA-ladder
Novabiochem	N_α-Fmoc-protected amino acids, 2-chlorotritylchloride resin, HBTU, HOBt, PyBOP, Castro's reagent
Operon	oligonucleotides
Oxoid	agar Nr.1, tryptone
Qiagen	oligonucleotides, QIAquick-spin PCR purification kit, Ni^{2+}-NTA-agarose, QIAexpress vector kit ATG, QIAEXII extraction kit, anti-His-antibody
Roth	Ethidium bromide, β-mercaptoethanol, acrylamide for SDS-PAGE, piperidine
Serva	Triton X-100, Visking dialysis tubes, APS
Sigma	EDTA, coenzyme A, N-acetylcysteamine, thiophenol, nucleotide pyrophosphatase, Basic and extension screening kits for crystallization
Stratagene	QuikChange^{TM} site-directed mutagenesis kit, PfuTurbo DNA polymerase

Vivascience AG	Vivaspin concentrators

3.2 Equipment

Device	**Manufacturer and Type**
Autoclave	Tuttnauer 5075 ELV
Bidestilled water supply	Seral Seralpur Pro90CN
Centrifugation	Heraeus Biofuge pico, Sorvall RC 26 plus, rotors SS34 und SLA3000, Sorvall RC 5B Plus, Kendro Megafuge 1.0R, Minifuge RF
DNA gel documentation	Cybertech CS1, thermoprinter Mitsubishi Video Copy processor
DNA sequence analyzer	Perkin-Elmer/ABI, ABI Prism 310 Genetic Analyzer
Electroporation-pulse control	Bio-Rad Gene Pulser II
FPLC system	Pharmacia FPLC-biotechnology FPLC-System 250: Gradient-programmer GP-250 Pump P-500 Uvicord optical device UV-1 (l = 280 nm) Uvicord control element UV-1 2-channel printer REC-102 Injection valve V-7 3-way-valve PSV-100 Fraction collector FRAC-100
French Press	SLM Aminco; French-Pressure Cell-Version 5.1; 20k Rapid-fill cell (40 mL)
HPLC-system	Agilent series 1100 HPLC-System with DAD-detection, vacuum degasser, quaternary pump, auto sampler and HP-chemstation software columns: Macherey & Nagel Nucleosil 250/3, pore diameter 120 Å, particle size 3 μm; Nucleodur-NH_2 250/3, pore diameter 100 Å, particle size 3 μm
MALDI-TOF	Per Septive Biosystems Voyager-DE RP BioSpectrometry, Bruker FLEX III
MS-MS sequencing	Applied Biosystems, API Qstar Pulsar I FT-ICR-MS
Peptide synthesizer	Advanced ChemTech APEX 396 synthesizer
Photometer	Pharmacia Biotech Ultraspec 3000
Shaker	New Brunswick Scientific Series 25 Incubator Shaker, New Brunswick Scientific Innova 4300 Incubator Shaker
Speed-Vac	Savant Speed Vac Concentrator,

	Uniequip Univapo 150
Thermocycler	Perkin-Elmer Thermal Cycler 480, Perkin Elmer Gene Amp PCR System 2400, Perkin Elmer Gene Amp PCR System 9700
Crystallization robot	Cartesian Microsys 4004
Crystallization plates	Hampton Research
Water bath	Infors Aquatron Shaker

3.3 Vector Systems

3.3.1 pQE60 and pQE61 vectors

The pQE60/61 vector system was used for cloning and overproduction of all Tyrocidine-NRPS-derived constructs (PCP-$C_{TycC5-6}$, PCP_{TycC5}, PCP_{TycC6}, C_{TycC6}, figure 4.1). The vector allows for purification of recombinant proteins by Ni-NTA chromatography by fusing a His_6-tag to the C-terminal end of the corresponding protein. The pQE60-vector carries two *lac*-operators in the promoter region. In the presence of a *lac*-repressor the gene can not be transcribed. Upon induction with IPTG, repression is abolished and gene transcription can occur. Therefore, this system allows for a defined start of protein production. The pQE61-vector is a derivative of pQE60, in which one *Bgl*II restriction site has been added directly after the *Nco*I site in the MCS.

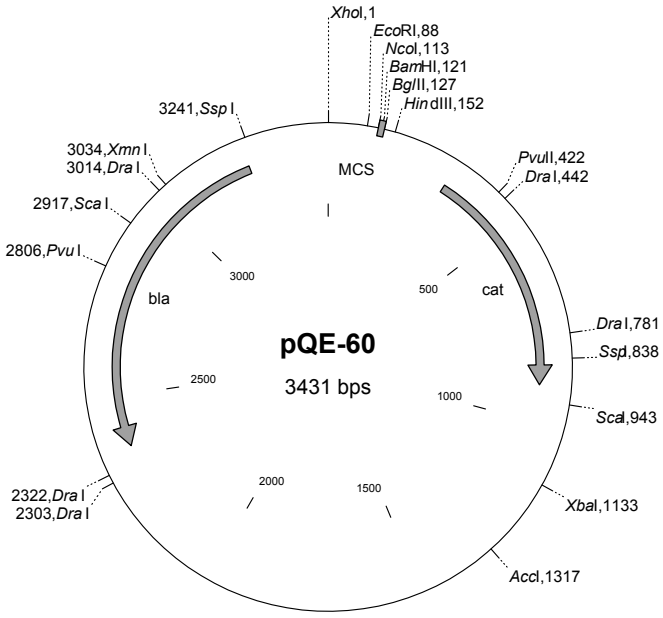


Figure 4.1: Map of the pQE60-vector.

The vector contains the following components:

origin of replication from *E. coli* (ColE1)

synthetic ribosomal binding site RBSII

T5-promotor from *E. coli*-phages

two *lac*-operator sequences for expression control by the *lac*-repressor

MCS with recognition sequences for: *Nco*I and *BamH*I

Stop codons in all three reading frames

codon sequence, which encodes a hexahistidine tag (C-terminal His_6-tag)

β-lactamase-gene *bla* for ampicillin resistance up to a final concentration of 100 µg/mL

two transcription terminators: t_0 of an λ-phage, T1 of the *rnnB*-operon from *E. coli*

3.3.2 pBAD202/D-TOPO vector

The pBAD202/D-TOPO vector system (Invitrogen) was used for cloning and expression of F-A-PCP_{LgrA1} and F-A-PCP-C-A-$PCP_{LgrA1-2}$. The vector is regulated by the araBAD-promoter (P_{BAD}) and is induced by arabinose. The His-patch thioredoxin leader (11.7 Da) increases translation efficiency and improves protein solubility. Removal of this thioredoxin fusion can be performed using EK (enterokinase) protease, which selectively recognizes the EK cleavage site. The vector also allows for Ni-NTA chromatography purification of recombinant proteins by fusion of a His_6-tag to the C-terminal end of the overexpressed protein.

The plasmid also contains the following components:

origin of replication from pUC plasmids

*rrn*B transcription terminator

codon sequence coding a V5 epitope

CAP-(cAMP binding protein) binding site for transcription enhancement by binding of the CAP-cAMP-complex

Kanamycin resistance

3.3.3 pREP4 helper plasmid

The pREP4 helper plasmid (Stratagene) is used in combination with the pQE vector system. Since it encodes for a *lac*-repressors, basal expression from genes inserted into pQE vectors is minimized, and transcription can sharply be induced by titrating the repressor with IPTG. It furthermore carries the *neo*-gene which provides a kanamycin-resistance up to 50 µg/mL. The plasmid also contains the origin of replication p15A from *E. coli* (compatible to pQE60 vectors).

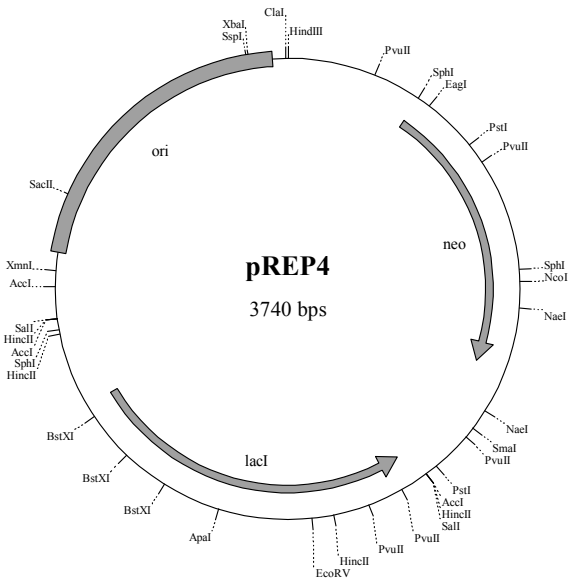


Figure 4.3: Map of the pREP4 plasmid.

3.4 Microorganisms

3.4.1 *E. coli* XL1-Blue

This strain was frequently used for cloning and sequencing purposes. The genotype is as follows: *rec*A1, *end*A11, *gyr*A96, *thi*-1, *hsd*R17, *sup*E44, *rel*A1, *lac*, F'(proAB^{+}, l*ac*I^{q}, *lac*ZDM15, Tn*10*(Tet^{I})).

3.4.2 *E. coli* TOP-10

E. coli Top 10 is another strain for cloning and sequencing purposes. The genotype is as follows: F- *mcr*A. (*mrr-hsd*RMS-*mcr*BC) 80*lacZ*.M15.*lac*X74 *deo*R *rec*A1 *ara*D139. (*ara-leu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G.

3.4.3 *E. coli* BL21 (DE 3)

The *E. coli* strain BL21(DE 3) with the genotype $F^- \ ompT[Ion] r_b^- m_b^-$ is used as a bacterial host for the expression of plasmid DNA. It is characterized by a lack of Ion protease and by a deficiency of OmpT protease, thereby significantly increasing protein stability. It further contains the IPTG-inducible T7 RNA polymerase gene, which is inserted in the chromosome after *lacZ* and the promoter *lac*uV5 on a λ-prophage. This is essential for the IPTG induction of genes under T7-promotor control.

3.4.4 *E. coli* BL21 (M15) [pREP4]

This strain lacks the T7-polymerase. It has the following genotype: *nals, strs, rifs, lac, ara, gal, mtl, F-*. It contains the pREP4 plasmid.

3.5 Media

E. coli strains were grown in LB-medium.

LB-medium:

16 g/L bactotrypton
10 g/L yeast-extract
5 g/L NaCl
pH 7.0, autoclaved

LB-medium based culture plates:

1.2% (w/v) of agar no.1 was added to the LB-media and heated at 121°C and 1,5 bar for 30 min (autoclave). Antibiotics were added after cooling down to ~50°C in the following standard concentrations: 100 µg/mL ampicillin (amp^{100}), 50 µg/mL kanamycin (kan^{50}). 34 µg/mL chloramphenicol (cm^{34}).

3.6 Buffers

HEPES-based buffers were used for protein purification and the assays, while a phosphate buffer was used for some of the F domain-related experiments. Their composition was as follows:

HEPES A:
50 mM HEPES
300 mM NaCl
pH7.0

HEPES B :
50 mM HEPES
300 mM NaCl
250 mM imidazole
pH 7.0

HEPES assay buffer:
50 mM HEPES
100 mM NaCl
pH 7.0

Phosphate buffer:
120 mM K_3PO_4
50 mM β-mercaptoethanol
pH 6.0

4. Methods

4.1 Construction of Recombinant Plasmids

Amplification of all DNA gene fragments was performed by polymerase chain reaction (PCR) with *Pfu* Turbo DNA polymerase (Stratagene) according to the manufacturer's protocol with chromosomal DNA. Purification of PCR-fragments was performed by the "QIAquick-spin PCR purification kit" in agreement with the manufacturer's manual (Qiagen). All constructs were analyzed by restriction digests combined with agarose gel electrophoresis and DNA-sequencing. Subsequent digestion and ligation steps were carried out in accordance with standard protocols [Sambrook 1989].

4.1.1 pBAD202-TOPO[F-A- PCP_{LgrA1}] and pBAD202-TOPO[F-A-PCP-C-A-$PCP_{LgrA1-2}$]

The pBAD Directional TOPO Expression Kit from Invitrogen (Paiseley, UK) was used to generate the desired pBAD202-TOPO expression plasmids. Cloning and preparation of the plasmids was performed in *E. coli* TOP10 (Invitrogen). The corresponding genes were PCR-amplified from chromosomal DNA of *Bacillus brevis* ATCC 8185. Primers used for the construction of pBAD202-[F-A-PCP_{LgrA1}] are as follows: 5′-CAC CGT GAG AAT ACT ATT CCT AAC AAC-3′, and 5′-TTG CTC CGT AAG CAG ACG-3′. For pBAD202-[F-A-PCP-C-A-$PCP_{LgrA1-2}$], 5′-CAC CGT GAG AAT ACT ATT CCT AAC AAC-3′, and 5′-GAA TTC GGA CGT GAC GAA TGG GGC-3′ were used.

4.1.2 pQE61[PCP_{TycC5}], pQE61[PCP_{TycC5}] and pQE61[C_{TycC6}]

The genes were amplified from chromosomal DNA *Bacillus brevis* (ATCC 8185) using degenerate primers introducing 5'-*Nco*I and 3'-*BamH*I restriction sites for ligation into the *Nco*I- and *BamHI*-linearized pQE61 vector (Qiagen). For pQE61[PCP_{TycC5}] 5'-TAA CCA TGG TTA GAT CTG AGT AGT TAG CGC CGC GC-3' and 5'-TTT GGA TCC CTA TGT CTC TTC GAT GAA CGC CGC CAG-3', while for pQE61[PCP_{TycC6}] 5'-AAA CCA TGG AAT ACG TGG CCC CGA GG-3' and 5'-AAA GGA TCC GAT GAA ATC GGC CAC CTT TTC G-3', and for pQE61[C_{TycC6}] 5'-TTT CCA TGG GAG GGA TGT CTT

CTC GAT CGA G-’3 and 5’-AAA GGA TCC AAG CAT GTC GAT CTC GCC C-3’ were used.

4.1.3 pQE61[PCP-$C_{TycC5-6}$] H224A andH224V mutants

The core 3 mutants H224A and H224V mutants were produced from the pQE61[PCP-$C_{TycC5-6}$] plasmid [Schoenafinger 2003] plasmids by site-directed mutagenesis (Stratagene kit) according to the manufacturer’s standard protocol. For H224A 5’-CTC TTT ACC GAC ATG CAT GCC AGC ATT TCC GAT GGC G-3’ and 5’-CGC CAT CGG AAA TGC TGG CAT GCA TGT CGG TAA AGA G-3’ primers were used, and for H224V 5’-CTC TTT ACC GAC ATG CAT GTC AGC ATT TCC GAT GGC G-3’ and 5’-CGC CAT CGG AAA TGC TGA CAT GCA TGT CGG TAA AGA G-3’ were applied.

After PCR the crude products were subjected to the *Dpn*I digestion, eliminating the template DNA due to its methylated DNA.

4.2 Construction of Expression Strains

As expression systems, either *E. coli* BL21 (M15) [pREP4] (C domain constructs) or BL21 (DE 3) (F domain constructs) strains were used. Since the electroporation method was to be used, the cells contained in 2 mL of expression strain culture (LB, $OD_{600\ nm}$ = 0.4 - 0.6) were harvested by centrifugation and washed seven times with sterile water at 4°C to remove salts, and finally resuspended in 40 µL. The according plasmids were dialyzed against water for 30 minutes at ambient temperature, before the electrocompetent recipient strains (40 µL cell suspension) were transformed with 1 µL plasmid solution (~0.1 ng/µL) by electroporation at 2.5 kV in cuvettes at 4 °C. Immediately after this procedure, 1 mL LB medium was added and the cells were incubated at 37 °C, 700 rpm for one hour. Then the culture was spread out on LB plates containing the appropriate antibiotics for selection, and incubated over night at 37 °C. The strains used for gene expression were derived from single colonies which were tested for the correct plasmid by sequencing thereof.

4.3 Protein Expression

For protein expression, overnight cultures of the expression strains were used to inoculate LB medium (1:100), supplied with antibiotics according to the resistances provided by the strains/plasmids. The inoculated medium was incubated at 37 °C until an $OD_{600\ nm}$ of ~0.5 was reached, upon which the temperature was lowered to 30 °C. After another 30 minutes, the expression inducing agents (IPTG, 0.1 mM for pQE-expression systems; 0.08% (w/v) arabinose for pBAD-systems) were added and the cells were harvested by centrifugation after another three hours.

The folate dehydrogenase FolD (*Methanosarcina barkeri)* was analogously produced from *E. coli* BL21 CodonPlus (DE3) pET24b+[folD] which was a generous gift from Dr. Bärbel Buchenau and Prof. Dr. R. Thauer, MPI Marburg [Buchenau 2004].

E. coli BL21 [pREP4-*gsp*] expression strains containing pQE60-TycA and pQE60-TycB1 plasmids were a generous gift from Dr. Martin Hahn, and the *holo*-proteins were analogously produced.

4.4 Protein Purification

The cell pellets were resuspended in HEPES buffer A and lysed using a French pressure cell™. The crude lysate was centrifuged at 17000 rpm for 40 minutes and the supernatant was subjected to Ni-NTA affinity chromatography by either FPLC or gravity flow columns at 4 °C. After loading, the sample was washed with HEPES buffer A until the flow-through was protein-free according to a Bradford test [Bradford 1976]. A linear gradient between 95% HEPES buffer A : 5 % HEPES buffer B and 5 % HEPES buffer A : 95 % HEPES buffer B was applied at a flow rate of 1 mL per minute (FPLC), or stepwise elution in percentual increments between 5 and 10 % HEPES buffer B was used for the gravity flow columns. The collected fractions were analyzed via SDS-PAGE [Laemmli 1970] and pooled accordingly. The combined fractions were then dialyzed against their 100-fold volume with the HEPES dialysis buffer (pH 7.0). Proteins were finally concentrated (10 – 100 μM) using Vivaspin columns, as determined by photometric measurements in accordance to the predicted extinction coefficients at 280 nm (Protean, DNA-Star software).

4.5 Synthesis of Aminoacyl- and Peptidyl-CoA Substrates

Peptidyl-CoA substrates were synthesized by automated solid phase peptide synthesis in 0.1 mmol scale using an Advanced ChemTech APEX 396 synthesizer. 2-chlorotrityl resin (IRIS biotech) and a standard Fmoc coupling strategy were used as described before [Belshaw 1999]. Monomers and coupling reagents were purchased from Novabiochem. Cleaved peptides were washed by preciptiation from 50 mL hexane and dried under reduced pressure.

After solubilization of the protected peptides and amino acids, repectively, with 4 ml of a mixture of THF:H_2O (1:1), 1.5 eq PyBOP, 1.5 eq CoA (trilithium salt), and 4 eq of K_2CO_3 were added and the reaction was stirred at ambient temperature for 2 h. Solvents were again evaporated under reduced pressure, prior to deprotection of the acid-labile protective groups by adding 4 ml TFA:H_2O:TIPS (95:2.5:2.5). The solution was stirred for 1 h at ambient temperature. The crude product was preciptitated from 30 ml cold diethylether and subjected to preparative reverse phase chromatography (Agilent-HPLC, C_{18}-column (Macherey-Nagel), linear gradient between 5 % MeCN and 95 % MeCN in water (all containing 0.1 % TFA) within 40 min). Automated fraction collection was triggered by UV absorption at 215 nm). The desired fractions were identified by MALDI-TOF, dried *in vacuo* and dissolved in Assay buffer (50 mM HEPES, 100 mM NaCl, pH 7.0) to a concentration of 5 mM. Purity was verified by LCMS. The substrates synthesized in this work are listed in section 5.2.3.

4.6 Synthesis of Transition State Analogs

Molecules resembling the putative transition state during peptide bond formation were synthesized according to a previously described synthetic route [Sellergren 2000]. The following figure 4.1 shows the synthetic route:

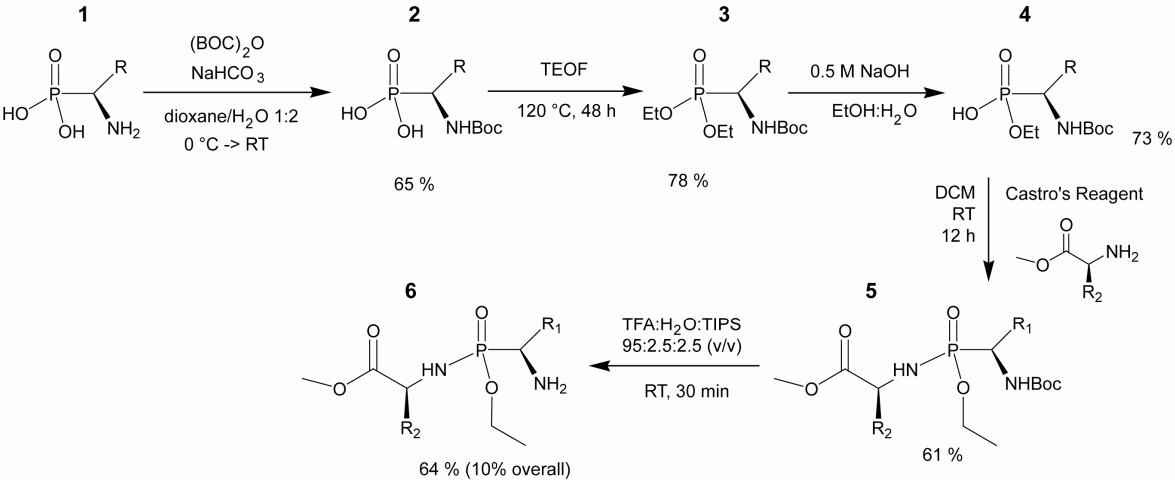


Figure 4.1: Synthetic route to transition state analogs (6) of the peptide bond forming reaction. R_1 and R_2 are methyl groups.

The single reactions are described below, numbers according to figure 4.1 were used to reference the molecules.

Synthesis of **2**:

To an ice-cole solution of **1** (250 mg, 2.00 mmol) in 20 mL of an aqueous 1 M solution of sodium carbonate, Boc_2O (2.20 g, 10.0 mmol) solved in 5 mL dioxane was added. It was stirred over night at ambient temperature, and washed twice with 20 mL diethylether each. The aqueous phase was acidified to pH 1 with HCl_{aq}, and the product was extracted three times with chloroform/isopropanol (3:1, 40 mL each). The united organic phases were dried over sodium sulfate, filtrated, and the solvents were removed *in vacuo*.

Analysis:

Yield:	296 mg (1.31 mmol, 65%).
DC:	R_f = 0,00 (chloroform/methanol 9:1).
^{1}H-NMR:	300 MHz, $CDCl_3$; δ/ppm = 10,77 (s, 2H, NH_2^+), 4,03 (q, 1H, $^3J_{HH}$ = 7,1 Hz, N-C*H*(CH_3)-P), 3,90 (s, 2H, O*H*), 1,44 (d, 3H, $^3J_{HH}$ = 7,1 Hz, N-CH(C*H*$_3$)-P), 1,36 (s, 9H, C(C*H*$_3$)$_3$).
^{13}C-NMR:	75 MHz, $CDCl_3$; δ/ppm = 146,7 (O-*C*(=O)-N), 80,5 ($(CH_3)_3C$-), 59,0 (N-*C*(CH_3)-P), 27,3 ($(H_3C)_3$C-), 15,5 (N-CH(*C*H_3)-P).
MS (ESI$^-$):	for $C_7H_{16}NO_5P$ (M-H$^+$) calc.: 224.2; obs.: 224.1.

Synthesis of **3:**

Compound **2** (296 mg, 1.31 mmol) was solved in TEOF (20.0 mL) and stirred for 72 h at 120 °C. TEOF was removed at 5 mbar and 45 °C, upon which a yellowish oil was obtained. It was solved in EtOAc (40 mL) and washed with water (50 mL each) twice. The organic phase was dried over sodium sulfate, filtrated, and the solvent removed *in vacuo*.

Analysis:

Yield: 290 mg (1.03 mmol, 78%).

DC: R_f = 0.29 (acetylethylester/hexane 4:1).

^{1}H-NMR: 300 MHz, $CDCl_3$; δ/ppm = 4,76 (d, 1H, $^3J_{HH}$ = 9,4 Hz, N*H*), 4,12 (m, 5H, CH_3CH_2-, -NH-C*H*(CH_3)-P-), 1,43 (s, 9H, (CH_3)$_3$-C-), 1,38 (d, 3H, $^3J_{HH}$ = 7,2 Hz, -NH-CH(C*H*$_3$)-P-), 1,31 (t, 6H, $^3J_{HH}$ = 7,0 Hz, -CH_2-C*H*$_3$).

^{13}C-NMR: 75 MHz, $CDCl_3$; δ/ppm = 155,0 (-*C*(O)O-), 80,1 ((CH_3)$_3$-*C*-), 62,8 (CH_3CH_2-), 62,5 (CH_3CH_2-), 43,7 (-NH-*C*H(CH_3)-P-), 28,4 ((CH_3)$_3$-*C*-), 16,6 (-NH-CH(*C*H_3)-P-), 16,6 (*C*H_3CH_2-).

^{31}P-NMR: 300 MHz, $CDCl_3$; δ/ppm = 27,6 (-*P*(O)(OEt)$_2$).

MS (ESI$^-$): for $C_{11}H_{24}NO_5P$ (M-H$^+$) calc.: 281.3; obs.: 281.0.

Synthesis of **4:**

Compound **3** (290 mg, 1.03 mmol) was solved in 22.5 mL ethanol, and 5.0 M $NaOH_{aq}$ was added until its final concentration reached 0.5 M. It was then stirred for 7 h at 60 °C. Ethanol was removed, and 20 mL water was added, before it was extracted with diethylether twice (25 mL each). The aqueous phase was acidified to pH 1 by adding HCl_{aq}, before the product was extracted with chloroform/isopropanol 3:1 three times (25 mL each). The united organic phases were dried over sodium sulfate, filtrated and the solvents removed.

Analysis:

Yield:	190 mg (0.75 mmol, 73%).
DC:	R_f = 0.00 (dichloromethane/methanol 9:1).
^{1}H-NMR:	300 MHz, $CDCl_3$; δ/ppm = 8,06 (s, 2H, -NH_2^+), 5,18 (s, 1H, -O*H*), 4,14 (m, 2H, H_3C-CH_2-), 4,04 (m, 1H, N-C*H*(CH_3)-P), 1,46 (s, 9H, (CH_3)$_3$C-), 1,33 (t, 3H, $^3J_{HH}$ = 6,7 Hz, H_3C-CH_2-), 1,21 (d, 3H, $^3J_{HH}$ = 6,1 Hz, N-CH(CH_3)-P).
^{31}P-NMR:	300 MHz, $CDCl_3$; δ/ppm = 28,0 (-*P*(O)(OH)OEt).
MS (ESI$^-$):	for $C_9H_{20}NO_5P$ (M-H$^+$) calc.: 252.1; obs.: 252.1.

Synthesis of **5**:

Substance **4** (95 mg, 373 µmol) was solved in dichloromethane (10.0 mL), cooled with an ice bath, before DIPEA (325 µL, 1.87 mmol), Castro's BOP Reagent (Benzotriazolyloxy-tris[dimethylamino]phosphonium hexafluorophosphate [Castro 1976], 412 mg, 932 µmol) and (*S*)-alanine-ethylester (143 mg, 932 µmol) were added and the reaction stirred for 24 h at RT. Solvents were partially removed until 3 mL remained. 10.0 mL of saturated $NaCl_{aq}$ were added and the product was extracted with acetylethylester (50.0 mL). It was washed twice with 1 M HCl_{aq} (20.0 mL each), saturated sodium hydrogencarbonate solution (20.0 mL each) and saturated $NaCl_{aq}$ (20.0 mL each). The organic phase was dried over sodium sulfate, filtrated and the solvents were removed. The crude product was further purified by silica gel column chromatography (chloroform/methanol 99:1).

Analysis:

Yield: 161 mg (0.458 mmol, 61%).

DC: R_f = 0.60 (dichloromethane/methanol 10:1).

^{1}H-NMR: 300 MHz, $CDCl_3$; δ/ppm = 4,77 (d, 1H, $^3J_{HH}$ = 9,3 Hz, P-N*H*-C), 4,11 (m, 6H, H_3C-C*H*$_2$-O-P, H_3C-C*H*$_2$-O-C(=O), N-C*H*(CH_3)-P, N-C*H*(CH_3)-C(=O)-), 1,42 (s, 9H, (*H*$_3$C)$_3$C-), 1,32 (m, 12H, *H*$_3$C-CH_2-O-P, *H*$_3$C-CH_2-O-C(=O), N-CH(C*H*$_3$)-P, N-CH(C*H*$_3$)-C(=O)-).

^{13}C-NMR: 75 MHz, $CDCl_3$; δ/ppm = 155,1 (d, $^3J_{CP}$ = 30,6 Hz, $(H_3C)_3$C-O-*C*(=O)-N), 80,1 ($(H_3C)_3C$-), 62,6 (dd, 2C, $^2J_{CP}$ = 83,1 Hz, $^5J_{CP}$ = 6,6 Hz, H_3C-*C*H_2-O-P, H_3C-*C*H_2-O-C(=O)), 43,7 (N-*C*H(CH_3)-C(=O)-), 41,6 (N-*C*H(CH_3)-P), 28,4 (($H_3C)_3$C-), 16,6 (N-CH(*C*H_3)-C(=O)-), 16,5 (N-CH(*C*H_3)-P), 16,4 (H_3C-CH_2-O-C(=O)), 16,3 (d, $^3J_{CP}$ = 109,4 Hz, H_3C-CH_2-O-P).

^{31}P-NMR: 300 MHz, $CDCl_3$; δ/ppm = 26,5 (-C(CH_3)-*P*O(OEt)NH-).

MS (ESI^-): for $C_{14}H_{29}N_2O_6P$ ($M\text{-}H^+$) calc.: 352.2; obs.: 352.2.

Synthesis of **6**:

Compound **5** (7.00 mg, 19.8 µmol) was solved in 2.50 mL of a mixture of TFA:H_2O:TIPS (2.38 mL:62.5 µL:62.5 µL) and stirred for 30 min at ambient temperature. TFA was removed under reduced pressure and the crude product was solved in toluene (10 mL), before the solvents were removed once more *in vacuo*. The crude product was purified by silica gel column chromatography (acetylethylester, then chloroform/ methanol/ triethylamine 99:1:1, and 95:5:1 at last).

Yield: 74 mg (0.29 mmol, 64%).

DC: R_f = 0.16 (chloroform/methanol 14:1).

RS

^{1}H-NMR: 300 MHz, $CDCl_3$; δ/ppm = 8,34 (d, 1H, $^3J_{HH}$ = 4,7 Hz, -P-N*H*-C-), 4,16 (m, 5H, $H_3C\mathit{H}_2OOC$-, $H_3C\mathit{H}_2OOP$-, H_2N-C*H*(CH_3)-P-), 3,88 (sextett, 1H, $^3J_{HH}$ = 6,8 Hz, -HN-C*H*(CH_3)-COO-), 1,80 (s, 2H, H_2N-), 1,59 (d, 3H, $^3J_{HH}$ = 6,8 Hz, -NH-C(CH_3)-COO-), 1,53 (d, 3H, $^3J_{HH}$ = 7,0 Hz, H_2N-CH(CH_3)-P-), 1,32 (q, 3H, $^3J_{HH}$ = 5,7 Hz, H_3CH_2OOC-), 1,32 (q, 3H, $^3J_{HH}$ = 8,3 Hz, H_3CH_2OOP-).

^{13}C-NMR: 75 MHz, $CDCl_3$; δ/ppm = 133,4 (d, 1C, $^3J_{CP}$ = 16,5 Hz, -*C*OO-), 63,8 (d, 1C, $^1J_{CP}$ = 170,0 Hz, H_2N-*C*H(CH_3)-P-), 62,9 (), 62,8 (), 62,6 (), 17,0 (), 16,9 (), 16,7 (d, 1C, $^1J_{CP}$ = 2,2 Hz,), 16,6 (d, 1C, $^1J_{CP}$ = 2,2 Hz,).

^{31}P-NMR: 300 MHz, $CDCl_3$; δ/ppm = 25,4 (-C(CH_3)-*P*O(OEt)NH-).

An overall yield of 10 % was achieved by these reactions. In parallel, the same synthesis was performed using the (*R*)-configured starting compound **1**, which led to an overall yield of 14 %**.** In both cases the absolute configuration of the stereocenter generated (**3**) could not be identified.

4.7 Synthesis of N^{10}-Formyl-Tetrahydrofolate

To produce N^{10}-formyl-tetrahydrofolate (N^{10}-fTHF), a chemoenzymatic synthesis in phosphate buffer was applied (see figure 5.4). Due to the instability of the product, is was always freshly prepared and used without further purification. The following compounds were united in a 200 µL quartz cuvette to a total volume of 200 µL, which was positioned in a photometer:

120	mM	K_3PO_4,
50	mM	β-Mercaptoethanol, pH 6.0
4	mM	NAD^+
2	mM	THF
4	mM (2eq)	HCHO

Hereby, the spontaneous formation of 5'-10'-methenyl-THF occurred. The regioselective oxidative ring-opening reaction that leads to the desired product was started by adding

2	µM	FolD

to the solution. The reaction progress was monitored, and indirectly quantified at 340 nm by the extinction coefficient of the stoichiometrically generated NADH ($\varepsilon_{NADH,\ 340nm}$ = 6.22 cm²/µmol) in relation to the measured OD. Slight background absorption was measured in a control reaction without FolD and subtracted accordingly. The overall yield was determined as 88 % and reached within ~10 seconds.

4.8 ATP/PP_i-Exchange Assays

The substrate specificity of A_{LgrA1} was determined by ATP/PP_i-exchange assays. The reversibility of the enzymatic aa-AMP-formation is used to incorporate (^{32}P-) radioactively labelled PP_i into ATP during the backward reaction. After a defined time, ATP is separated from the reaction mixture by adsorption onto charcoal, and the amount of radioactivity detected allows for a relative comparison of the reaction rates generated by the A domain in presence of different amino acid substrates.

The reaction mixtures (100 µL total volume) were composed as follows:

50	pM	F-A-PCP_{LgrA}
1	mM	amino acid
4	mM	ATP
0.2	mM	PP_i
4	mM	$MgCl_2$
0.15	µCi	(^{32}P)-PP_i

The separate reactions were started by addition of the amino acids tested (valine, isoleucine, leucine, and formylvaline) and incubated for 10 minutes at 37 °C. The reactions were stopped by adding 500 µL of a quenching solution composed of 100 mM sodium pyrophosphate, 560 mM perchloric acid, and 1.2 % (w/v) charcoal Norit A. The sample was vortexed and the charcoal separated by centrifugation for 1 minute at 13000 rpm. The resulting pellet was washed twice with destilled water and finally resuspended in it (500 µL each). After adding the sample to 3.5 mL of scintillation liquid (Rotiszint Eco Plus) in a scintillation vial, the radioactivity was measured in a scintillation counting device.

4.9 F Domain Assays

In order to produce the holo-enzymes, *apo*-F-A-PCP_{LgrA1} and *apo*-F-A-PCP-C-A-$PCP_{LgrA1-2}$ as obtained after protein purification (30 µM each) were incubated with 100 µM CoA, 1 µM Sfp, and 10 mM $MgCl_2$ for 30 min at 25 °C directly prior to subsequent assays without intermediate purification.

The *holo*-enzymes (25 µM) generated *in situ* were separately incubated with 100 µM ATP, 100 µM amino acid (valine, isoleucine or leucine for F-A-PCP_{LgrA1}; valine and glycine for F-A-PCP-C-A-$PCP_{LgrA1-2}$) and 112,5 µM of either N^{10}-fTHF prepared *in situ* as described above, or N^5-fTHF, respectively, in HEPES assay buffer (50 mM HEPES, 100 mM NaCl, pH 7.0). Samples with a total volume of 200 µL were kept at 25 °C and incubated for 2 h.

When non-cognate amino acids were tested for formylation with F-A-PCP_{LgrA1}, the *apo*-enyzme was provided with 100 µM of the according aminoacyl-CoA in the Sfp-dependent priming reaction instead of CoA. Consequently, no amino acids were furthermore added and the reaction was conducted otherwise identically.

Reactions were stopped and the proteins were precipitated by the addition of 1 mL 10 % (w/v) TCA. Samples were centrifuged at 4 °C and the pellet was washed three times with 500 µL Et2O:EtOH (3:1 (v/v)). Solvents were evaporated at 37 °C *in vacuo*. Thioester cleavage was performed by adding 100 µL KOH (0.1M) to the dry pellets and incubating at 70 °C for 15 min. 1 mL MeOH was added and samples were kept at -20 °C for 14 h. Precipitated KOH was removed by centrifugation and the supernatants were dried by evaporation at 45 °C and 10 mbar in a speed-vac system. The resulting pellets were dissolved in 30 µL 95 % acetonitrile$_{aq}$ containing 2mM NEt_3. For the time dependent formylation assays, identical procedures were used and samples were taken after 0, 1, 4, 10, and 100 minutes and immediately quenched by adding them to 1 mL 10% (w/v) TCA.

4.10 C Domain Assays

4.10.1 Cyclization Assays with PCP-$C_{TycC5-6}$ and the H224A and H224V Mutants

Apo- PCP-$C_{TycC5-6}$ (75 µM) was incubated with 2 µM Sfp in the presence of 10 mM $MgCl_2$ and 125 µM peptidyl-CoA in a total volume of 50 µL at 25 °C for 2 h. Reactions were stopped by adding 20 µL 4 % TFA, and directly subjected to LCMS analysis.

Identical procedures were applied, when the H224A and H225V mutants were tested for cyclization of the hexapeptidyl substrate DPhe-Pro-Phe-DPhe-Asn-Gln-CoA. In the inhibitory assays, 1 mM of the *R,S*-configured synthesized transition state analog was present in the assays.

4.10.2 *In trans* Experiments with PCP_{TycC5} and C_{TycC6}

The *apo*-PCP_{TycC5} (100 µM) was incubated with DPhe-Pro-Phe-DPhe-Asn-Gln-CoA (125 µM), in the presence of 0.5 µM Sfp and 10 mM $MgCl_2$ in HEPES assay buffer for 20 minutes at 25 °C. In parallel, 100 µM C_{TycC6} was incubated with the synthesized transition state analog (*R,S*-configuration) or other small molecules (see figure 5.28C) to be tested for inhibition. The concentration of the inhibitory molecules was adjusted to 100 µM in both pots, before the hexapeptidyl-*holo*-PCP_{TycC5} and C_{TycC6} were mixed (60 µM: 1 µM) to initiate the reaction. Time dependent samples were taken from the reaction mixture after 1, 5, 10, 20, 60 and 120 minutes, stopped by adding 4% (v/v) TFA and analysed by HPLC-MS. In additional (otherwise identical) experiments, several other commercially available (Aldrich) small tetrahedral or reactant-resembling molecules were used instead of the synthesised phosphonamides.

4.10.3 Diketopiperazine Formation Assay

The produced *holo*-TycA (4 µM) and *holo*-TycB1 (2 µM) enzymes were incubated with 1 mM of the *S,S*-configured phosphonamide transition state analog in HEPES assay buffer adjusted to pH 8.0. After 15 minutes, their cognate amino acids phenylalanine and proline (200 µM each) were added along with 10 mM $MgCl_2$, and the reaction was started by addition of ATP (10 mM). Time dependent samples were taken from the reaction mixture after 1, 5, 10, 20, 60 and 120 minutes and the reactions were stopped by adding 4% (v/v) TFA prior to LCMS analysis.

4.11 Analysis with LCMS

The assays were analyzed using high performance liquid chromatography (HPLC) combined with an electrospray-injection (ESI) mass spectrometer (MS), or "LCMS" for short (Agilent).

For the C domain assays, a reversed-phase chromatography was applied. A C_{18}-column (125/2 Nucleodur 100-5, Macherey-Nagel) was used, and acetonitrile/water supplemented with 0.1 % (v/v) TFA was chosen as solvent system. A standard volume of 40 µL of the assay samples was injected, and a linear gradient between 5 and 60 % acetonitrile within 20 minutes was applied for compound separation. The eluted substances were time-dependently detected by a UV-photometer set to 215 nm, and a fraction thereof submitted to online mass analysis. The mass spectrometer was used in the positive mode. When the flow rate was set to 0.4 mL/minute, peptidic analytes exhibited retention times between 8 and 20 minutes. This method was furthermore used to identify the CoA-substrates produced in this word.

For the F domain assays however, the standard reversed-phase setup was found inappropriate (poor retention times), and a normal-phase gradient was applied in the context of a stationary amino-phase (125/2 Nucleosil 100-5 NH_2, Macherey-Nagel). Water and acetonitrile were supplemented with 2 mM triethylamine, and the starting conditions were 95 % acetonitrile and 5 % water. Within 30 minutes, a linear gradient was applied to 20 % acetonitrile and 80 % water at a flow rate of 0.3 mL/minute. The products were detected by the mass spectrometer set to negative single ion detection mode.

4.12 Crystallization of *apo*-PCP-$C_{TycC5-6}$

Initial crystallization screens were performed at 18 °C using a Cartesian Microsys 4004 cystallization robot. Crystals were identified from one condition (1.6 M $(NH_4)_2SO_4$; 0.1 M MES (pH 6.5); 10% dioxane). Crystals suitable for crystallographic analysis were obtained within 3 days by Stefan A. Samel from *hanging drop* setups (Hampton Research) with a protein concentration of 7.5 mg/ml and a reservoir solution containing 1.6 M ammonium sulfate, 0.04 M MES (pH 6.5) and 4% dioxane.

5. Results

Amide bonds are ubiquitously found in nonribosomal secondary metabolites. Even though the vast majority of chemical bonds formed to build up these products are amide/peptide bonds, little is known about the catalytic mechanisms involved in their generation. Principally, three types of nonribosomal domains catalyze the formation of amide bonds: The condensation (C), thioesterase (TE), and formylation (F) domains. The most important contributors in NRPS are the condensation domains as they are generally needed for every elongation step during the build-up of the linear peptide product or the linear precursor for subsequent macrocyclization. So far, no coherent model for the C-domain catalyzed reaction mechanism is available although two somewhat contradictory explanations have been proposed in the past. The TE domains either catalyze the hydrolytic release or the macrocyclization of the products and are usually situated at or associated with the very C-terminal part of an NRPS machinery. Their reaction mechanisms are much more thoroughly understood, and one single amide bond is generated in case of a macrocyclization involving an amino-group nucleophile (macrolactamization). The formylation domains were hypothetic altogether at the beginning of this work.

Two different classes of amide bond forming enzymes were studied: The formylation and condensation domains of nonribosomal peptide synthetases. Firstly, the identification and characterization of the F domain from the linear gramicidin NRPS will be presented. Secondly, mechanistic and structural studies addressing the mode of catalysis of C domains and their interactions with PCP domains are described.

5.1 The Formylation Domain

Unlike ribosomal peptides isolated from bacteria, the nonribosomal peptide products do not generally carry a formyl unit on their N-terminal α-amino group. However, a few cases are known in which N_α-formylations are present: Linear gramicidin (see chapter 1.2) produced by *Bacillus brevis* ATCC 8185, and the anabaenopeptilides from *Anabaena* strain 90 (figure 5.1). Anabaenopeptilides are branched depsipeptides with a six-membered ring as central structural motif. Besides N_α-formylation, N-methylation, chlorination and O-methylation, they carry the unusual building blocks 3-amino-6-hydroxy-2-piperidone which is connected to threonine *via* an amide bond, and homo-tyrosine. The corresponding biosynthetic NRPS machinery (see appendix) is located in the 30 kb *adp* cluster and consists of three synthetases (AdpA, AdpB, and AdpD) which comprise 24 domains organized in 7 modules, while AdpC is a putative halogenase.

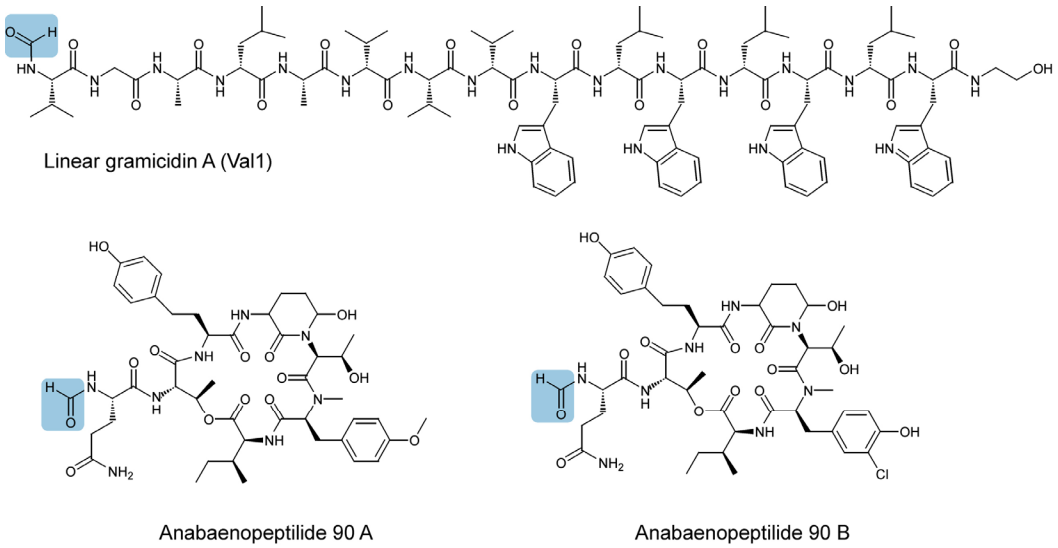


Figure 5.1: Nonribosomal peptides carrying a formyl group on their α-amino group.

Sequence analyses of the corresponding NRPS genes (*lgr*A-E and *adp*A-F) revealed the presence of a novel N-terminal domain type which was hypothesized to be responsible for the N_α-formylation found in the products, and it was therefore called formylation (F) domain. The following subsections cover the experiments carried out to investigate the F domain from the linear gramicidin system (F_{LgrA}).

5.1.1 Selected Constructs

Since the F domain is located at the N-terminus of the linear gramicidin NRPS system, the constructs chosen were derived from LgrA: F-A-PCP$_{LgrA1}$, F-A-PCP-C-A-PCP$_{LgrA1-2}$ (figure 5.2).

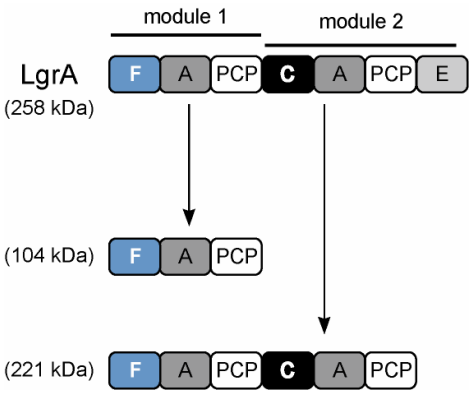


Figure 5.2: Protein constructs derived from the LgrA: F-A-PCP$_{LgrA1}$ and F-A-PCP-C-A-PCP$_{LgrA1-2}$.

With F-A-PCP$_{LgrA1}$, the substrate specificity for the formyl transferase reaction can be addressed *in cis* – allowing both artificial and natural (A domain mediated) loading of acceptor substrates onto the PCP domain's cofactor Ppan. The necessity of a formylated starter unit for the subsequent NRPS reactions can be tested by using F-A-PCP-C-A-PCP$_{LgrA1-2}$: Does the linear gramicidin NRP synthesis only proceed with the C$_{LgrA2}$-domain-mediated condensation reaction when the first building block is formylated?

5.1.2 Cloning, Expression, and Purification

On the DNA-level, the selected constructs were generated by PCR from chromosomal DNA of *Bacillus brevis* ATCC 8185. The gene fragments obtained were cloned into pBAD202/D-TOPO vectors using a one-step cloning strategy without ligase and restriction enzymes. This expression system appends an N-terminal His-patch thioredoxin domain (11.7 kDa) to the recombinant proteins, which facilitates solubility and translation efficiency, and allows for Ni-NTA purification of the expressed proteins. *E. coli* TOP10 cells were used for cloning steps, and *E. coli* BL21 was used for subsequent protein production. All strains were grown in LB medium, and expression was induced by addition of arabinose. Cells were harvested by centrifugation and disrupted using a French™

pressure cell prior to Ni-NTA chromatography of the soluble fraction of the crude lysate. HEPES-based buffers (pH 7.0) were used containing 0 to 250 mM IPTG (in stepwise increments of 50 mM) for washing and elution of the isolated recombinant proteins (figure 5.3).

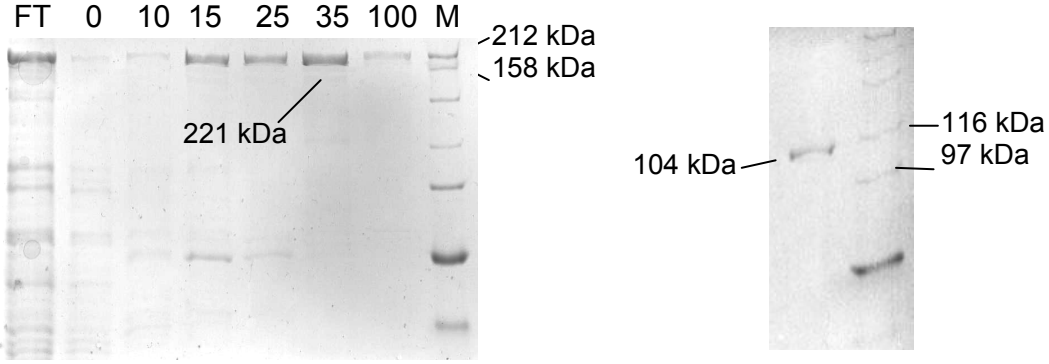


Figure 5.3: PAGE analysis of the recombinant proteins after Ni-NTA purification. On the left hand side the gel for F-A-PCP-C-A-PCP$_{LgrA1-2}$ (221 kDa) is presented. Pure protein is eluted at 35 % (v/v) buffer B. On the right hand side, the 104 kDa F-A-PCP$_{LgrA1}$ elution (100 %(v/v) buffer B) fraction is displayed.

5.1.3 N^{10}-Formyl-Tetrahydrofolate Synthesis

The generation of formylated methionine as the starter unit for ribosomal translation in bacteria is a prominent formylation reaction in nature. Methionyl-tRNA-formyltransferases (FMTs) catalyse the transfer of the formyl group onto the N_α-amino group of Met-tRNA^{fMet} by consumption of the cofactor N^{10}-formyl-tetrahydrofolate (N^{10}-fTHF). Sequence alignments of the F domains from *Bacillus brevis* ATCC 8185 (F_{LgrA}) and *Anabaena* Strain 90 (F_{AdpA}) with several such FMTs showed that both classes of enzymes contain the so-called "SLLP" motif [Schracke 2005]. This conserved sequence motif is suggested to be characteristic for the binding of the N^{10}-fTHF cofactor used by FMTs. It was therefore hypothesized that N^{10}-fTHF might also be the formyl group donor in F-domain-mediated reactions investigated here.

To generate N^{10}-fTHF, a chemoenzymatic approach was used in a one-pot synthesis. Initially, tetrahydrofolate is incubated with formaldehyde in aqueous, phosphate-buffered solution (pH 6.0, RT), which leads to the formation 5',10'-methenyl-tetrahydrofolate. This reaction intermediate serves as the substrate for the added folate dehydrogenase FolD [Buchenau 2004] that catalyzes oxidative, regioselective ring-opening giving rise to N^{10}-fTHF (Figure 5.4).

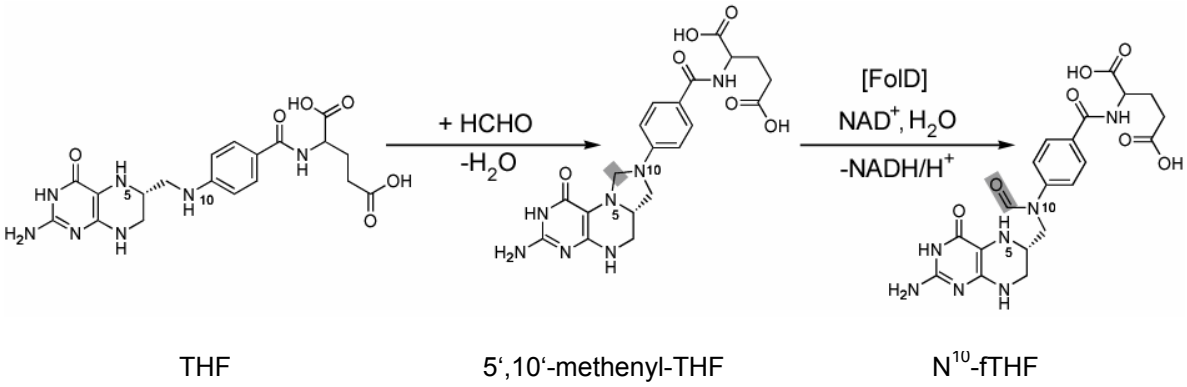


Figure 5.4: Reaction scheme of the N^{10}-fTHF synthesis. In the presence of formyldehyde, THF spontaneously forms 5',10'-methenyl-THF which is then regioselectively oxidized by FolD under consumption of NAD^+, leading to the desired product N^{10}-fTHF.

Since NAD^+ is stoichiometrically converted into NADH in this enzymatic redox reaction, the increase of light absorption at 340 nm can be monitored and used to indirectly quantify the reaction outcome. The reaction was therefore carried out in quartz cuvettes. Due to the instability of N^{10}-fTHF, the cofactor was produced directly prior to its use and without further purification steps. Thus, it was important to minimize side products and other non-desired impurities. In several rounds of optimization (Figure 5.5), the amount of formaldehyde could be drastically reduced, which is essential since formaldehyde is detrimental to enzymatic reactions (FolD and F domain constructs in assays) – yet necessary to produce 5',10'-methenyl-tetrahydrofolate *in situ*.

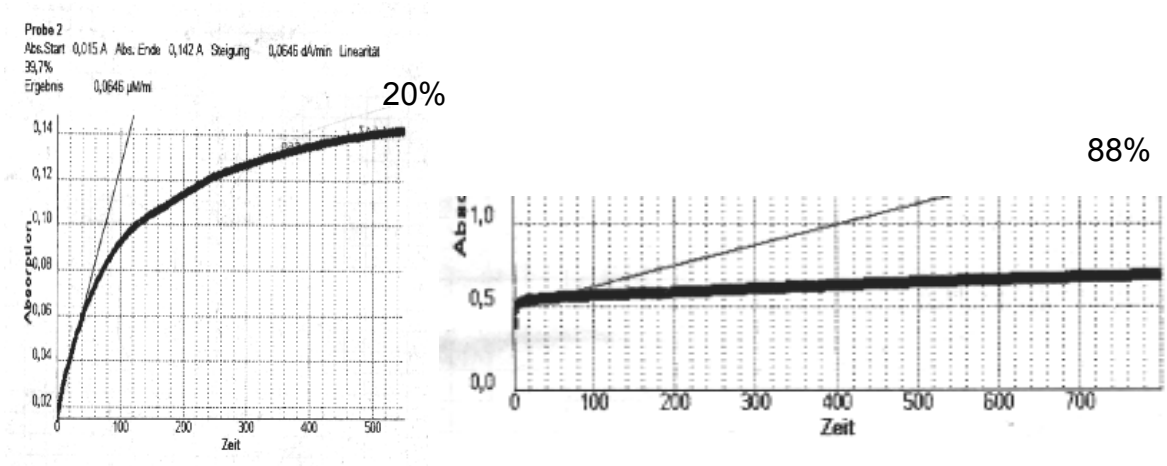


Figure 5.5: Time dependent formation of NADH during N^{10}-fTHF synthesis monitored photometrically. With the initial reaction setup (50 equivalents of formaldehyde), only ~20 % yield was obtained after 10 minutes, as depicted on the left side. After optimization of the reaction (2 equivalents of formyldehyde) an 88 % yield was achieved after ~10 seconds (right).

5.1.4 The Specificity of the Adenylation Domain A_{LgrA1}

Understanding the timing, directionality, and reaction hierarchy in enzymatic machineries as complex as NRPSs is essential when the role of a newly discovered functional domain is to be assessed. Since the exact formyl acceptor substrates of the F domain were unknown, the knowledge of the substrate specificity of the A domain in LgrA1 helps rule out several possibilities. If the A domain accepts the unmodified amino acids found at position 1 in the peptide sequence of linear gramicidin (valine, leucine, isoleucine) but not their formylated counterparts, then the formyltransferase activity is likely to be dependent on the presence of these substrates bound to PCP_{LgrA1}. If, on the other hand, the formylated amino acids were substrates for the A domain, then an *in trans* formylation of valine, leucine and isoleucine would have to precede the activation reactions.

To address this matter, ATP/PP_i-exchange assays were carried out using F-A-PCP_{LgrA1}. In these assays, the reversibility of the enzymatic aa-AMP-formation is used to incorporate (^{32}P-) radioactively labelled PP_i into ATP during the backward reaction. After a defined time, ATP is separated from the reaction mixture by adsorption onto charcoal, and the amount of radioactivity detected allows for a relative comparison of the reaction rates generated by the A domain in presence of different amino acid substrates. When an amino acid is readily accepted the reactions are fast whereas substrates that are not accepted by the A domain lead to very slow reaction rates, and therefore less measured radioactivity in the ATP. The following figure 5.6 summarizes the results of these experiments:

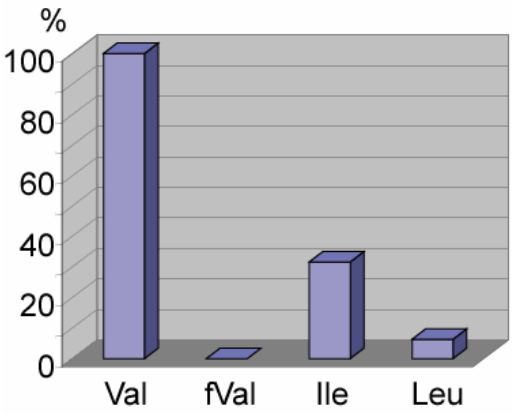


Figure 5.6: The A_{LgrA1} domain's specificity as determined by ATP/PP_i-exchange experiments. When the turnover for valine is defined as 100 %, isoleucine retains 31.7 % and leucine 6.3 % thereof while no reaction was observed when formylvaline was used.

5.1.5 Formyltransferase Assays using F-A-PCP$_{LgrA1}$

To test for potential formylation activity *in vitro*, *apo*-F-A-PCP was enzymatically converted into the corresponding *holo*-form using the phosphopantetheinyl-transferase Sfp and CoA. In a second step, the naturally occurring starter amino acids Val, Leu, and Ile were separately added along with ATP in 4-fold excess over F-A-PCP to allow for A-domain-mediated activation and loading of these formyl acceptor substrates (Figure 5.7). Formylation activity was triggered by the addition of ~ 4.5 equivalents freshly prepared N^{10}-formyl-tetrahydrofolate. Since the formylated reaction products and educts remain covalently bound to the enzymatic template, basic hydrolysis with potassium hydroxide to cleave the thioester bond between Ppan and the amino acid moieties was used to facilitate subsequent analysis.

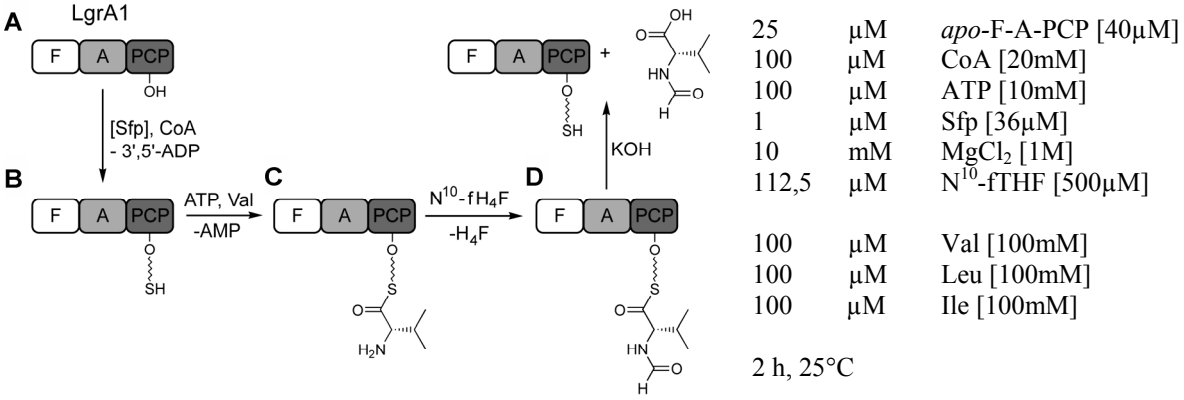


Figure 5.7: Scheme of the formylation assays using F-A-PCP$_{LgrA1}$. In the first step, the *apo*-enzyme is converted to *holo* **(A)**, then ATP and either valine, leucine, or isoleucine is supplied **(B)**, and the formylation reaction **(C)** is triggered by addition of the cofactor N^{10}-fTHF. Products can be detected from the solution after basic cleavage from the enzyme **(D)**.

Valine was quantitatively formylated, whereas the reactions with leucine and isoleucine did not show complete turnovers (figure 5.8). To quantify the corresponding ratios, a simple signal integration of these single-ion mass traces is inappropriate due to the expected differences in the ionization efficiencies between the formylated and non-formylated species. Thus, external standards for the amino acid and the formylated amino acid were used to establish a correction factor (Figure 5.8). Under the HPLC-MS conditions applied, the formylated species exhibit a 15-fold greater ionization efficiency.

With this correction factor, the ratio of formylated to non-formylated leucine and isoleucine can be determined as 20:1 – or 5 % reaction yield in both cases.

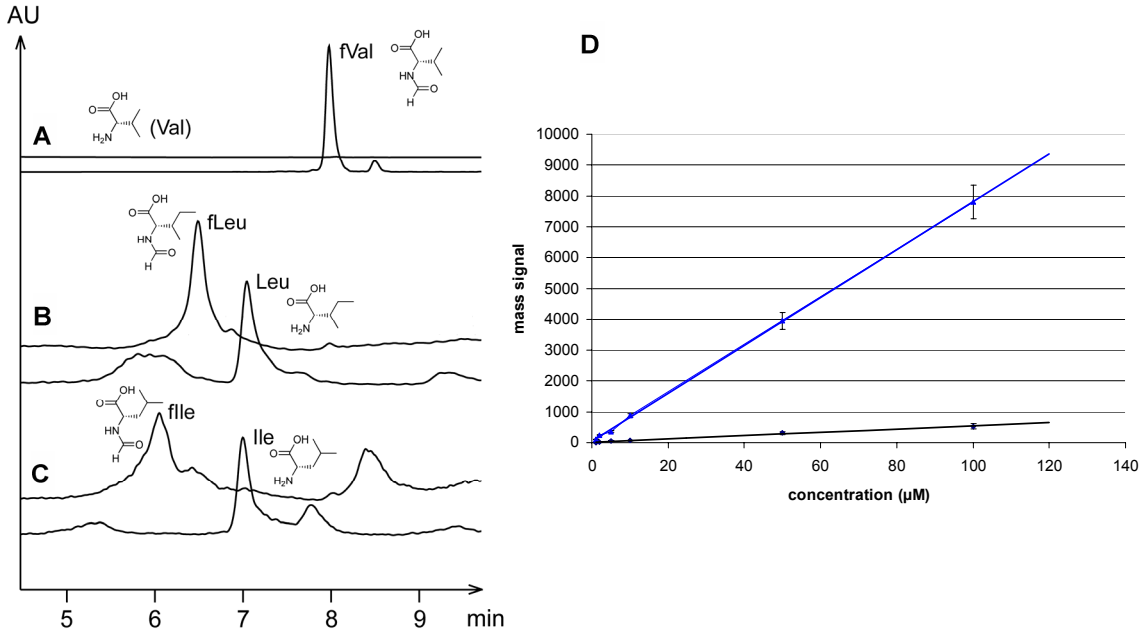


Figure 5.8: HPLC traces of the assay analysis and determination of the ionization efficiencies of formylated and non-formylated valine. **(A)** Valine is quantitatively converted to formyl-valine while leucine **(B)** and isoleucine **(C)** are only formylated in 5 % yields, when the empirically determined correction factor for the ionization efficiencies is applied. **(D)** Mass signal intensities for formyl-valine (black) and valine (blue) plotted as a function of the concentration. The slope factor difference is 15 and was used as a correction factor when quantifying the products in **(B)** and **(C)** by intergration of the UV-chromatogram signals.

5.1.6 Formyl Acceptor Substrate Specificity

Even though a preference of the formylation reaction towards PCP-bound valine as acceptor substrate was indicated in previous experiments (5.1.5), a variety of other amino acid substrates was tested in analogous assays. However, the A domain could not be used for enzymatic loading because of its specificity towards the naturally occurring (branched aliphatic) amino acids valine, leucine, and isoleucine. To circumvent this problem, aminoacyl-CoAs were synthesized (see also 5.2.3) and used in the modification reaction of the *apo*-enzyme (Figure 5.9). As a result, the corresponding aminoacyl-Ppan moieties were directly transferred onto the carrier protein. The assay conditions and work-up were otherwise identical, yet no formylated amino acids could be detected apart from the naturally occurring ones.

Substrate	Val	Ile	Leu	Ala	Thr	Asn	Glu	Lys	Tyr	Phe
Formylation yield [%]	100	~5	~5	0	0	0	0	0	0	0

Figure 5.9: Amino acyl substrate acceptance in the F-A-PCP$_{LgrA1}$ catalyzed formylation reaction. Only the branched aliphatic amino acids Val, Ile, and Leu were accepted.

5.1.7 Formyl Donor Substrate Specificity

With respect to the reaction intermediate 5',10'-methenyl-tetrahydrofolate during the formyl donor substrate synthesis (Figure 5.5), the question arose whether N^5-fTHF might also be a suitable substrate for the F domain. This compound is commercially available, more stable, and its use would avoid side products and impurities that arise from the *in situ* synthesis of N^{10}-fTHF, as described in 5.1.3. In two parallel assays, *apo*-F-A-PCP was used with valine as acceptor substrate, and equal amounts of N^5-fTHF and N^{10}-fTHF, respectively, as donor substrates. The assay procedure was otherwise identical to the one described in 5.1.5, even though a scale-up to a reaction volume of 400 µL was necessary to obtain sufficient sample volumes for a time course comparison. Samples were extracted after 0, 1, 4, 10, and 100 minutes. The amount of produced formyl-valine was quantified by comparison of the integrals of the corresponding MS signals. In both reactions, the maximum was already reached after 10 minutes (Figure 5.10) and remained unchanged even after 100 minutes. However, the reaction with N^{10}-fTHF was found to be faster – despite the expected negative effects of the side products of the *in situ* synthesis of this cofactor (5.1.3). When defining the reaction speed by the elapsed time at which half of the maximum turnover is reached, the reaction with N^{10}-fTHF is roughly 18-fold faster than the one with N^5-fTHF.

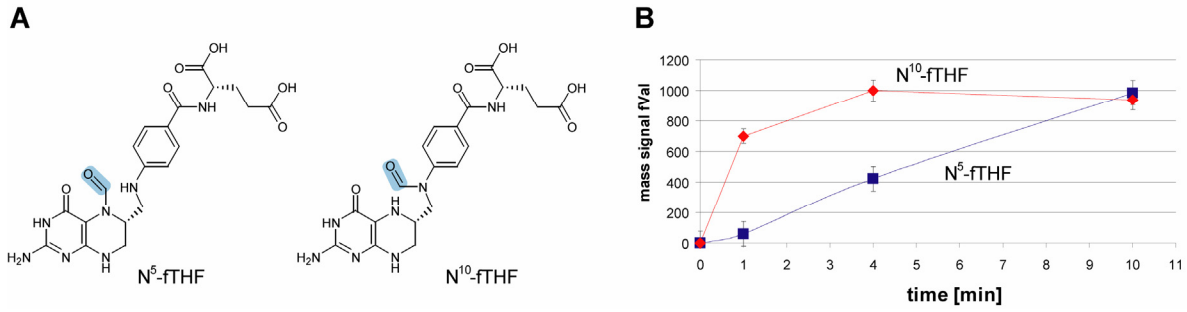


Figure 5.10: Comparison of the N^5- and N^{10}-fTHF cofactors. **(A)** The position of the formyl groups in the structure formulae. **(B)** Time courses of formyl-valine production with N^5-fTHF (blue) and N^{10}-fTHF (red). The same maxima are reached after 10 minutes, yet the reaction with the N^{10}-type cofactor is ~18-fold faster.

5.1.8 The Necessity of a Formylated Starter Unit in Linear Gramicidin Biosynthesis

All known isoforms of linear gramicidin carry an N-terminal formyl group. Even though the domain responsible for the N-formylation had been identified, it was unclear whether the NRPS machinery would intrinsically guarantee that only such formylated starter amino acids were further processed or if the relative reaction speeds of the single reactions merely statistically lead to practically exclusively formylated products. If the latter were true then *in vitro* experiments should be able to show non-formylated products, yet if the former were true some sort of regulatory mechanism would prevent further reactions with non-formylated starter units. To address this matter, *apo*-F-A-PCP-C-A-PCP$_{LgrA1-2}$ was used. After enzymatic *apo/holo*-conversion with Sfp, the cognate amino acids valine (for A$_{LgrA1}$) and glycine (for A$_{LgrA2}$) were supplied along with ATP (Figure 5.11).

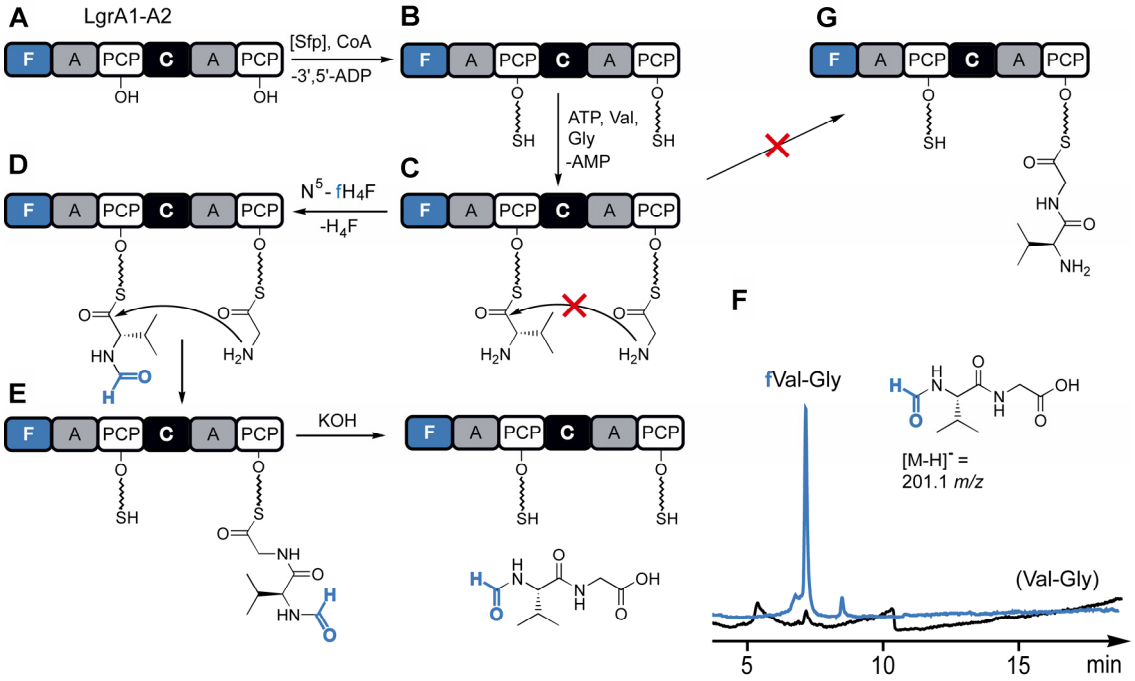


Figure 5.11: Scheme of the assays performed with F-A-PCP-C-A-PCP$_{LgrA1-2}$. **(A)** The apo-enzyme is converted into its holo-form by the Ppan transferase Sfp. **(B)** The cognate amino acids valine and glycine are provided for the A-domain-mediated loading reaction. **(C)** In the primed enzyme, the C domain catalyzed condensation reaction **(G)** does not occur, unless **(D)** the formyl donor is present, in which case the formylation of the first amino acid valine precedes the condensation reaction **(E)**. The product is hydrolyzed off the template prior to detection. **(F)** The chromatograms of the mass signals reveal that fVal-Gly (blue) is produced while the unformylated Val-Gly (black) is not formed.

As a result, aminoacyl-*holo*-enzyme was formed. From here on, two possible routes could be followed by the NRPS machinery: Firstly, the condensation reaction occurs – leading to the Val-Gly dipeptide. Secondly, no dipeptide is formed as long as the starter building

block is not formylated, but formyl-Val-Gly is produced when the formyldonor cofactor is supplied. Thus, two separate reactions were carried out: One with all components needed for the (putative) formation of formyl-Val-Gly, and a control which lacks the formyldonor. The reaction, work-up and detection conditions were analogous to those used for F-A-PCP_{LgrA1}, as described in 5.1.5.

The results show that no unformylated condensation product was detected in the control while formylated dipeptide was produced when the formyldonor was present.

5.2 The Condensation Domain

This section covers experimental studies, both mechanistic and structural, that address the nature of the peptide bond formation catalyzed by condensation (C) domains. Even though C domains are found in every NRPS system and their sequences and sequence similarities are well known, no coherent model for their catalysis on a molecular level exists. The studies presented here were aimed at shedding light on the mechanism of catalysis and substrate recognition.

5.2.1 Selected Constructs

The Tyrocidine NRPS gene cluster (*tyc*) from *Bacillus brevis* (ATCC 8185) was used as a model system, and the selected enzyme constructs were thus derived from the synthetases TycA, TycB, and TycC (Figure 5.12).

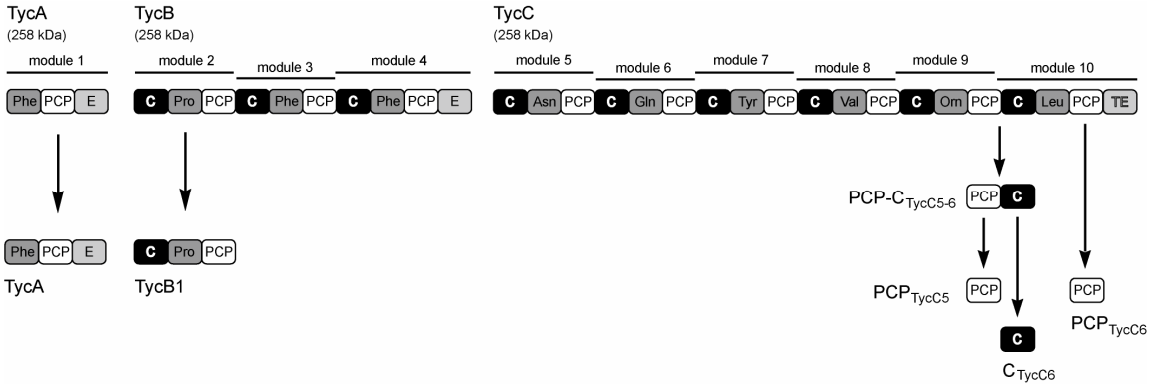


Figure 5.12: Selected constructs derived from the tyrocidine synthetases TycA, TycB, and TycC.

Both the first (TycB1) and last (TycC6) C domains were chosen in these selected constructs as they fulfill roles as divergent as possible within this system: The first C domain condenses two aminoacyl building blocks and needs to act *in trans* – communicating with TycA. The last C domain, on the other hand, acts *in cis* and condenses a nonapeptide with an amino acid. The latter was used for both substrate specificity studies and crystallization experiments while the former was chosen for inhibitory assays along with TycA – taking advantage of the DKP formation side reaction as a marker for activity.

As for C_{TycC6}, the isolated neighboring PCP domains were generated to allow for *in trans* experiments.

5.2.2 Cloning, Expression, and Purification

The gene fragments for PCP_{TycC5}, C_{TycC6}, PCP_{TycC6}, and PCP-$C_{TycC5-6}$‡ were amplified by PCR using degenerate primers and chromosomal DNA of *Bacillus brevis* ATCC 8185. The primers were designed so that subsequent cloning into the vector pQE61 was possible *via* *Nco*I and *Bam*HI restriction sites. The vector appends a C-terminal hexahistidine tag. *E. coli* TOP10 cells were used for cloning steps, and *E. coli* M15 [pREP4] was used for subsequent protein production. All strains were grown in LB medium, and expression was induced by addition of 0.1 mM IPTG at 28 °C. Cells were harvested by centrifugation and disrupted using a French™ pressure cell prior to Ni-NTA chromatography of the soluble fraction of the crude lysate. HEPES-based buffers (pH 7.0) were used containing 0 to 250 mM IPTG for washing and elution of the isolated recombinant *apo*-proteins (Figure 5.13).

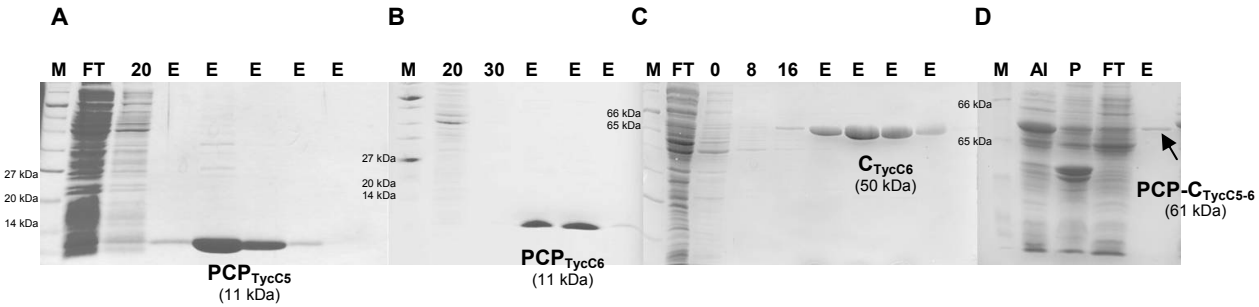


Figure 5.13: PAGE analysis of the selected constructs derived from TycC. **(A)** PCP_{TycC5}, **(B)** PCP_{TycC6}, **(C)** C_{TycC6}, and **(D)** PCP-$C_{TycC5-6}$. M = protein marker, FT = flow-through, lanes marked with numbers define the (v/v)-percentage of buffer B containing 250 mM imidazole used for washing steps, E = elution fractions with 100 % buffer B, AI = crude lysate after induction, P = pellet after centrifugation.

E. coli BL21 [pREP4-*gsp*] expression strains containing pQE60-TycA and pQE60-TycB1 plasmids were a generous gift from Dr. Martin Hahn, and the *holo*-proteins were analogously produced. Figure 5.14 shows the PAGE analysis of the production.

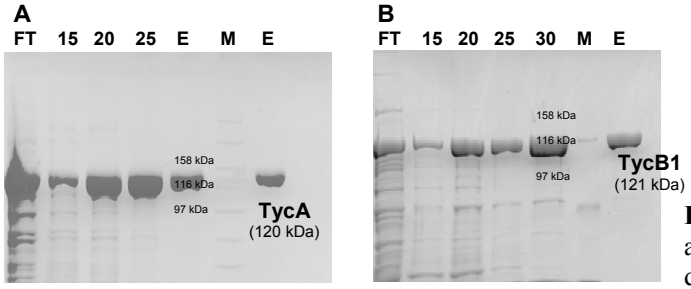


Figure 5.14: Production of *holo*-TycA **(A)** and *holo*-TycB1 **(B)**. Lanes are named as described in figure legend 5.13

‡ This construct had already been made during the author's diploma thesis work.

5.2.3 Synthesis of Aminoacyl/Peptidyl-CoAs

A common technique for modifying *apo*-carrier proteins is the enzymatic transfer of acylated phosphopantetheinyl groups originating from their corresponding CoA-thioesters [Sieber 2003]. As a result, the acylated *holo*-forms are obtained in a single step, and the enzyme of interest can thereby easily be provided with a broad variety of acyl and peptidyl substrates.

Both aminoacyl and peptidyl substrates were produced in this work. As for peptidyl substrates, the oligopeptides were synthesized on solid support (chlorotrityl resin), and an acid labile protective group strategy was applied. The protected peptides were cleaved from the resin, precipitated from hexane and dried *in vacuo*. For the subsequent coupling reaction, the crude oligopeptides or protected amino acids, respectively, were activated with PyBOP and coupled to CoA. After deprotection with TFA the products were precipitated from cold diethylether and purified reverse phase HPLC. The eluted fractions were analyzed by MALDI-TOF MS spectrometry and the solvents were removed from the desired products by lyophilization. The substrates used in this work are listed in the following table:

Aminoacyl-CoA substrates	Val	Ile	Leu	Ala	Thr	Asn	Glu	Lys	Tyr	Phe
Peptidyl-CoA substrates	P1: DPhe-Pro-Phe-DPhe-Asn-Gln									
	AcP1: Ac-DPhe-Pro-Phe-DPhe-Asn-Gln									
	P2: LPhe-Pro-Phe-DPhe-Asn-Gln									
	P3: DAla-Pro-Phe-DPhe-Asn-Gln									
	P4: LAla-Pro-Phe-DPhe-Asn-Gln									
	P5: DPhe-Pro-Phe-DPhe-Asn-Ala									
	P6: DPhe-Pro-Phe-DPhe-Ala-Ala-Asn-Gln									
	P7: DPhe-Pro-Phe-Asn-Gln									

5.2.4 Elongation/Cyclization Assays with PCP-$C_{TycC5-6}$

In the diploma thesis work of the author, PCP-$C_{TycC5-6}$ had already been used in assays aiming at an *in vitro* elongation of aminoacyl or peptidyl substrates by an additional amino acid. The general strategy for these experiments is to enzymatically load substrates onto the PCP domain of the bidomain construct and supply potential acceptor substrates *in trans* (Figure 5.14).

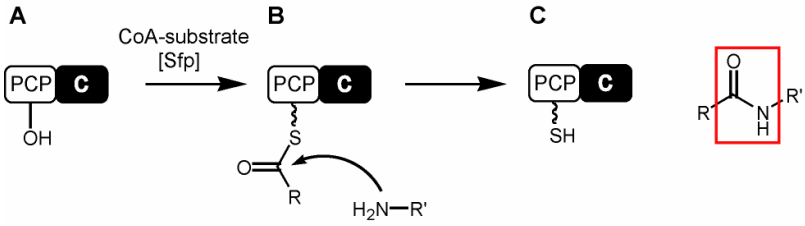


Figure 5.14: General assay strategy for *in trans* elongation experiments with PCP-$C_{TycC5-6}$. **(A)** The *apo*-enzyme is loaded with its donor substrate using a CoA derivative and Sfp. **(B)** An amino nucleophile attacks the thioester in a C domain catalyzed reaction. **(C)** After the single-turnover reaction, a product with the newly generated amide bond (red) has been generated, and the enzyme remains in its *holo*-state.

When the hexapeptidyl substrate DPhe-Pro-Phe-DPhe-Asn-Gln was used as donor substrate, neither the free peptide acid (hydrolysis) was found as the main product nor traces of the heptapeptide (elongation) were detectable in the presence of leucinyl-SNAc acceptor *in trans* (Figure 5.15). However, a species with a mass reduced by 18 compared to the hydrolysis product was the main product – with a ratio of 20:1 over the uncatalyzed reaction. It was hypothesized that this was the cyclic counterpart of the hexapeptide.

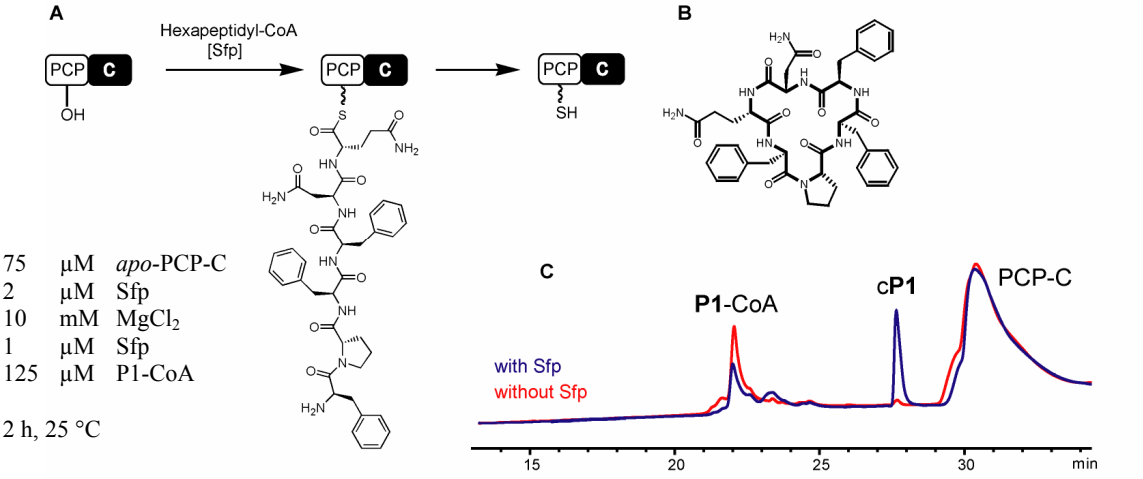


Figure 5.15: The hexapeptidyl-cyclization reaction with *apo*-PCP-$C_{TycC5-6}$. **(A)** Reaction scheme and conditions. **(B)** Chemical formula of the cyclic product. **(C)** UV-chromatogram of the HPLC analysis. In presence of Sfp, P1 is cyclized to cP1 (blue) while there is only the slight background cyclization observable when Sfp is omitted in a control reaction (red).

With regard to potential nucleophiles within the hexapeptidyl donor substrate that might be involved in this cyclization, no suitable nucleophiles are present aside from the N-terminal amino group. Thus, a head-to-tail connectivity was suspected and to be indirectly proven by using an N-terminally protected variant of the substrate in an analogous experiment. The substrate Ac-DPhe-Pro-Phe-DPhe-Asn-Gln-CoA was therefore synthesized and used under otherwise identical conditions. In this case, no corresponding species with a mass reduced by 18 was found, and again only traces of hydrolysis product

were detected. However, now that the cyclization reaction was suppressed, still no elongated heptapeptide was produced in the presence of the "natural" acceptor substrate leucin or the SNAc variant thereof. The cyclic hexapeptide was isolated and subjected to MS^n analysis. One fragment was identified that proves the expected head-to-tail connectivity as it contains both Gln and DPhe-Pro-DPhe – which is only possible if a covalent bond between the C- and N-termini exists in the sample material (Figure 5.16).

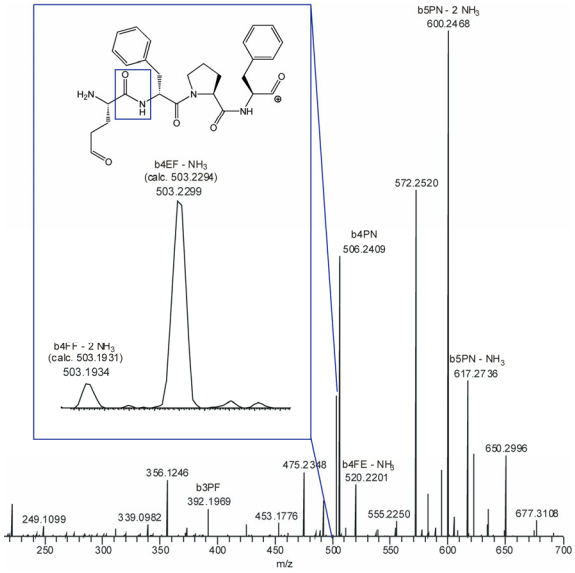


Figure 5.16: MS^n analysis of the cyclic hexapeptide. The fragment highlighted belongs to a Phe-Pro-Phe-Gln sequence that can only occur when a bond between Phe1 and Gln6 exists – proving the head-to-tail connectivity.

In further experiments, other peptidyl substrates were used to investigate the substrate tolerance of the cyclization reaction with PCP-$C_{TycC5\text{-}6}$. With lengths between 5 and 8 amino acids, the type and the absolute configuration of the terminal amino acids were altered in comparison to DPhe-Pro-Phe-DPhe-Asn-Gln. The PCP-C bidomain only cyclized substrates which contained D-configured N-termini (P1, P3, P7; Table 5.1). No hydrolysis product was cleaved off the enzyme except in the case of the octapeptidyl substrate P6. The C-terminal L-Gln cannot be substituted by L-Ala in these reactions. Whenever a C domain dependent cyclization was observed, autocatalytic formation of the cyclic product also occurred with relative amounts of 5 to 17 % as determined by peak integration of the UV/Vis-chromatograms.

5.2.5 Core 3 Mutants in Cyclization Experiments

Since the cyclization reaction observed was dependent on the presence of peptidyl-*holo*-PCP-$C_{TycC5-6}$, the question remained whether the C domain was performing active catalysis.

In previous studies, the second histidine of sequence motif "HHxxxDG" (core 3) in the C domain had been suggested to be essential for condensational catalysis [Bergendahl 2002]. To investigate this matter in terms of cyclization activity, the according H224 residue was mutated by site-directed mutagenesis to alanine and valine, respectively, giving rise to two mutants of PCP-$C_{TycC5-6}$: H224A and H224V. The production and purification conditions of these proteins were analogous to the ones used for the wild-type protein (see 5.2.4). Both mutant proteins were soluble and showed the typical α-helical fold CD-spectra, as depicted below (Figure 5.17).

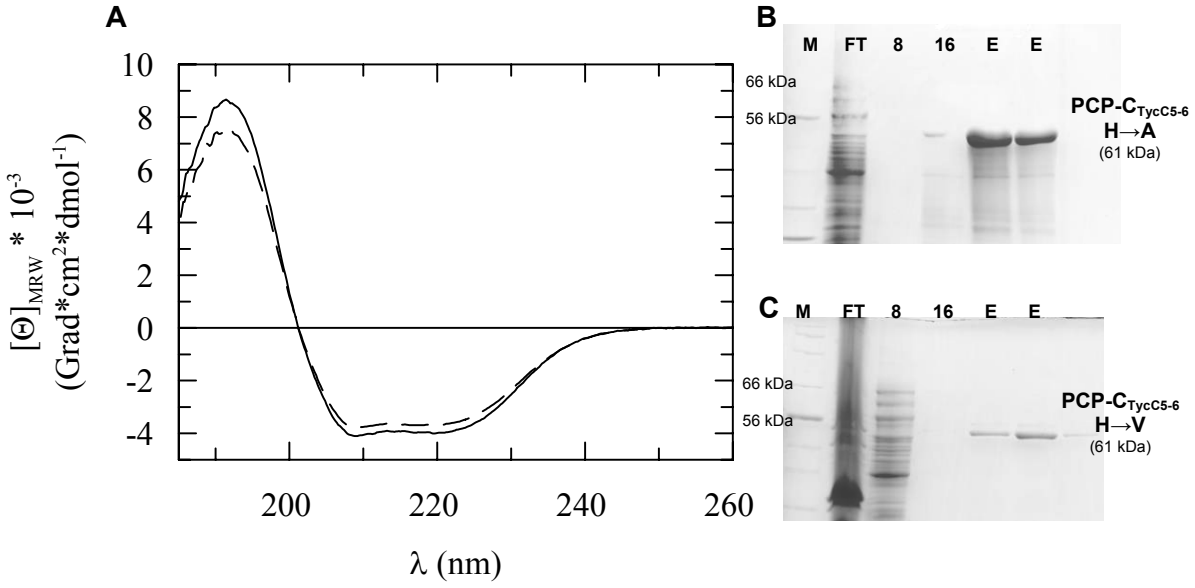


Figure 5.17: CD spectra and PAGE analysis of the core 3 mutants PCP-$C_{TycC5-6}$ H→A and H→V. **(A)** The α-helical spectrum of the wild-type enzyme (dashed line) superimposed with the H→A mutant spectrum (solid line). The H→A **(B)** and H→V mutant **(C)** as seen by SDS-PAGE. Lanes are named as described in figure legend 5.13.

Cyclization experiments with the hexapeptide DPhe-Pro-Phe-DPhe-Asn-Gln lead to no increased formation of cyclization product over the background when the H224A and H224V mutants were used (figure 5. 18), under otherwise unchanged conditions (see 5.15).

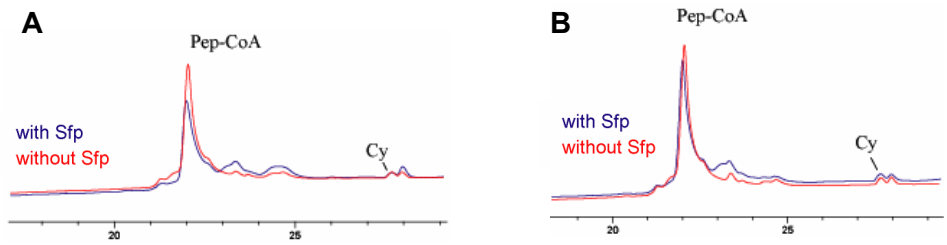


Figure 5.18: HPLC UV-chromatograms of the cyclization experiments with the H224A **(A)** and H224V **(B)** mutants of PCP-$C_{TycC5-6}$. In both cases, the cyclization activity is not increased (blue) compared to the controls where Sfp was omitted (red).

5.2.6 Crystallization of PCP-$C_{TycC5-6}$

The *apo*-enzyme of PCP-$C_{TycC5-6}$ was screened for suitable crystallization conditions using the hanging drop method in a 2 µL scale. Two protein concentrations were used (7.5 and 15 mg/mL), and 1 µL of these samples was mixed with 1 µL of crystallization buffer each. These buffers were taken from the commercially available Sigma basic® and Sigma extension® screening kits. The reservoirs on the crystallization plates were filled with 400 µL of these buffers prior to sealing the setups and storing them at 18 °C. Inspection with a light microscope revealed the formation of numerous small crystals (figure 5.19) formed under the following conditions after seven days: 1.6 M ammonium sulfate, 0.1 M MES (pH 6.5) and 10 % 1,4-dioxane. At this point, a cooperation with the structural biochemistry group of Prof. Dr. L.-O. Essen at the biochemistry department of the Philipps-University in Marburg, Germany, was initiated. The project was thus handed over to Dipl.-Chem. Stefan A. Samel who carried out further optimization studies and managed to drastically improve the nucleation ratio and crystal size as required for proper X-ray diffraction. The following pictures show sample crystals before and after the optimization process (figure 5.19).

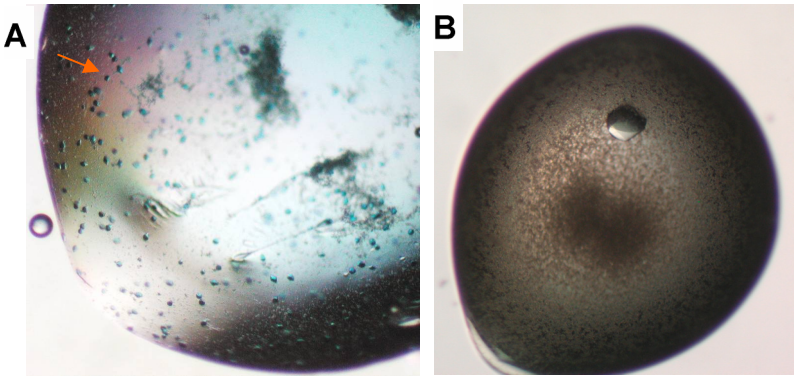


Figure 5.19: Pictures of the crystallization studies with *apo*-PCP-$C_{TycC5-6}$ before **(A)** and after optimization **(B)**.

Stefan A. Samel furthermore produced and crystallized variants of the enzyme in which methionine was substituted by selenomethionine. The structure was then solved at a resolution of 1.8 Å by multiple anomalous diffraction experiments (PDB accession number 2JGP) [Samel 2007].

5.2.7 The Structure of the Bidomain Enzyme

In this subchapter, the crystal structure of the *apo*-PCP-$C_{TycC5\text{-}6}$ enzyme is described since it substantially contributes to the understanding of C domains and nonribosomal domain interactions. Moreover, further inhibitory experiments were inspired by it, and the C domain's catalytic mechanism is discussed here on the basis of these insights.

Apo-PCP-$C_{TycC5\text{-}6}$ is a distinct bidomain enzyme in which the PCP (M1-T82) and C (V101-L522) domains both exhibit their own foldings. The PCP domain measures 20 Å x 24 Å x 30 Å, roughly, whereas the C domain can be fitted into a cuboid with edge lengths of 65 Å x 50 Å x 40 Å. Both domains are connected *via* an 18-residue linker (A83-P100) running along the PCP-C domain interface (figure 5.20).

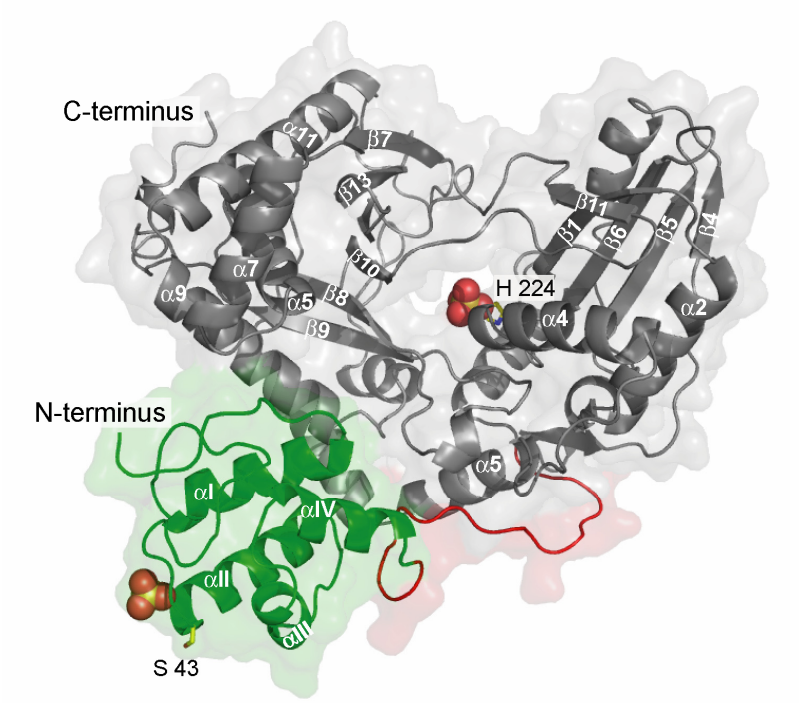


Figure 5.20: Overall structure of the bidomain enzyme *apo*-PCP-$C_{TycC5\text{-}6}$. The PCP domain (green) is connected to the C domain (grey) by an 18-residue linker (red). Two sulfate ions (sphere models) are seen in the crystal structure, both located closely next to conserved residues (H244 and S43).

5.2.7.1 The Condensation Domain's Substructure

In analogy to a previously described structure for the stand-alone C domain VibH [Keating 2002], C_{TycC6} consists of two mainly separated and structurally similar subdomains: one N-terminal (V101-S268) and one C-terminal (A269-L522) subdomain. These are arranged in a V-shaped fashion and belong to the chloramphenicol-acetyltransferase (CAT) fold (figure 5.20). Only two major contact sites between the two CAT-like subdomains exist, giving rise to a large canyon-like active site groove. The first is located at the floor of the active site canyon and comprises the loop β8-β9 (T359-V374) which forms mostly H-bond interactions with the N-terminal subdomain. The second region is strand β11 (N438-F465) that stretches from the C-terminal to the N-terminal subdomain where it extends the four-stranded β-sheet (β1-β6-β5-β4). This region spans like a bridge over the active site canyon and appears to be rather flexible due to high thermal B-factors as compared to the rest of the domain.

A superposition of C_{TycC6} with VibH (PDB code 1L5A, pairwise sequence identity 19 %) shows structural similarity with an overall rmsd of 1.58 Å for 197 C_α-carbons (figure 5.21). However, only the C-terminal CAT-like subdomains are fitted properly in this superposition as the corresponding N-terminal CAT-like subdomain of C_{TycC6} is off-rotated by about 12° compared to the VibH subdomain pairing. The hinge-like region – around which this swivelling motion is centered – corresponds to S268 (in VibH: S174) in the short connection between helices α5 and α6. As a consequence, the structural comparisons between the corresponding subdomains yield significantly lower rmsd values of 1.35 Å for 149 C_α-carbons of the C-terminal and 1.70 Å for 116 C_α-carbons for the N-terminal CAT-like subdomains.

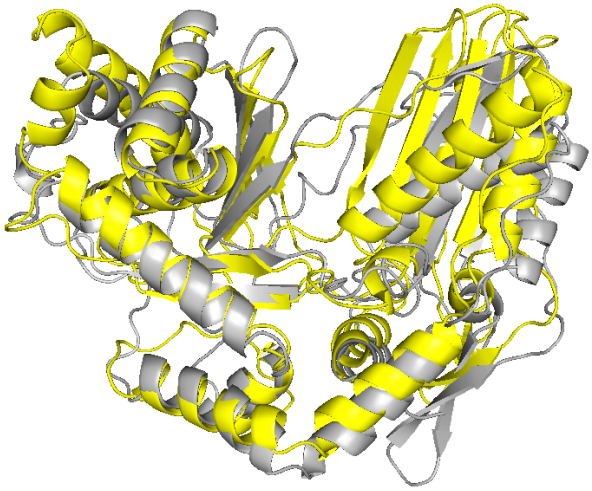


Figure 5.21: Superposition of C_{TycC6} (grey) with VibH (yellow). The N-terminal subdomains are off-rotated by approximately 12° (right) while the C-terminal subdomains (left) superimpose properly.

In the crystal structure, a buffer-derived sulfate ion is bound in proximity of the supposed active site at the canyon floor (figure 5.22). It forms a salt bridge to the residue H224 (Nε2 2.87), an H-bond to the backbone amide nitrogen of G229 (2.68 Å) and is further stabilized by the dipole moment of helix 4. The H224 is the second histidine of the core 3 motif of C domains which was previously suggested to play a critical role for C domains' catalytic activity [Bergendahl 2002], and was found to be necessary for hexapeptidyl cyclization in this work (see 5.2.4 and 5.2.5). The well-ordered binding of a sulfate ion in this very location was interpreted as a possible structural surrogate for the sp^3-type reaction intermediate during the peptide-bond-forming reaction. The binding of this sp^3-species by H224 would accelerate the peptide bond formation by transition state stabilization rather than acid-base catalysis (see discussion).

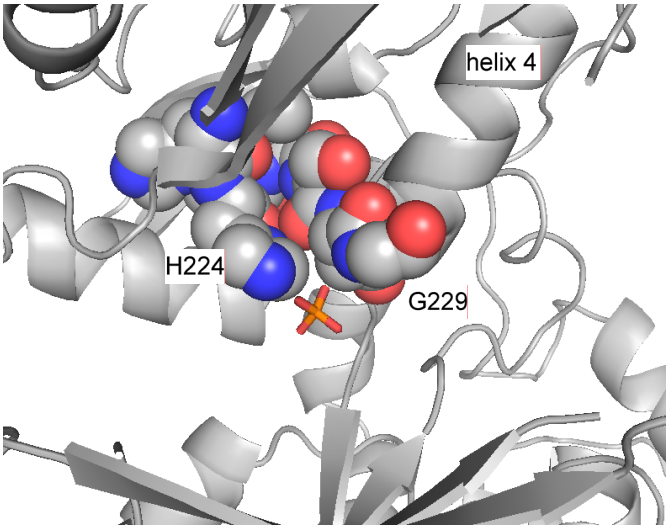


Figure 5.22: A sulfate ion (stick model) is bound to active site by interactions with H224, G229 and helix 4. The residues of the core 3 motif are shown as spheres.

5.2.7.2 The PCP Domain's Substructure

The structure of the *apo*-PCP domain (figure 5.20) closely corresponds to the A/H-state [Koglin 2006]. This state was found in the PCP domain of the third module of TycC as an intermediary conformational state that is present in both, the Ppan-modified (*holo*, "H") and unmodified PCP domain (*apo*, "A"). A superposition with the A/H-like state of the PCP domain of module 3 of TycC gives a structural similarity of 1.67 Å for 56 C_α–positions. In the *holo*-form of PCP domains, the A/H-state is in conformational equilibrium with the H-state, which becomes increasingly stabilized by interactions with other *holo*-PCP recognizing domains like the editing thioesterase II [Koglin 2006]. This H-state

diverges significantly from the observed conformation of PCP_{TycC5}, as 78 C_α-positions superimpose with an rmsd of only 5.5 Å, mostly along helices α1 and α4 (18 $C_\alpha \equiv 2.3$ Å).

The sulfate ion next to the PCP-domain (figure 5.20) forms salt bridges to the conserved residues H42 (Nδ 3.11 A) and R45 (Nε 2.92 A; Nη 2.89 A) and is additionally stabilized by the dipole moment of helix αII, which points with its N-terminus onto this sulfate. Obviously, this sulfate anion occupies the supposed position of the phosphate group of the Ppan arm when the latter is esterified to S43 of the PCP domain.

5.2.7.3 The Linker Region

The short peptide sequences between the distinctly folded non-ribosomal domains are highly variable and called linkers. In the bidomain crystal structure here, this linker region consists of 18 amino acid residues (A83-P100). It does not show any secondary structure motifs, even though some interactions are found: hydrophobic interactions formed by F88 with the residues W261 and F265 of the C domain. Furthermore, several residues of the linker are involved in an intricate H-bond network with both, the PCP- and C domain (figure 5.23). Not only does N86 interact with T82 (3.15 Å) and the carboxylic group of I79 (3.18 Å), but its Oδ-atom also forms an H-bond with the amide nitrogen of F88 (2.91 Å). Likewise, the side chain of D257 interacts with the amide nitrogen of V93 (2.91 Å) whereas the indole amine function of W261 is H-bonded to the amide oxygen of I90.

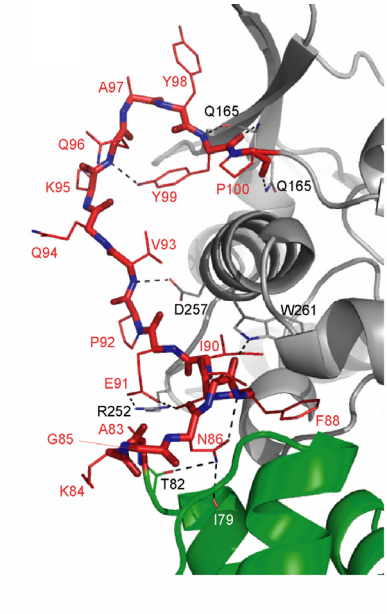


Figure 5.23: A view of the linker region (red). Polar interactions with the PCP (green) and C domain (grey) are indicated by dashed lines.

5.2.8 Design and Synthesis of Potential Inhibitors for the Condensation Domain

Both the tightly bound sulfate ion in the catalytic site of the C domain and the addition/elimination nature of the peptide-bond-forming reactions in NRPS led to the idea of synthesizing a stable transition state analog for inhibitory and crystallographic studies. The main assumption was that the tetrahedral sulfate ion likely resembles the transition state of the condensation reaction just after the initial nucleophilic attack of the α-amino group onto the thioester carbon (figure 5.24). In this step, the planar thioester functionality is formally converted into the sp^3-configured zwitterionic intermediate. Thereby, a stereocenter is temporarily generated which – upon elimination of the thiol moiety in the subsequent step – is later destroyed again. For designing a stable transition state analog, the high reactivity of the naturally occurring sp^3-species had to be reduced while maintaining as much similarity as possible. Therefore, the central carbon atom was replaced by phosphorous, and the oxyanion was consequently changed to a P=O unit.

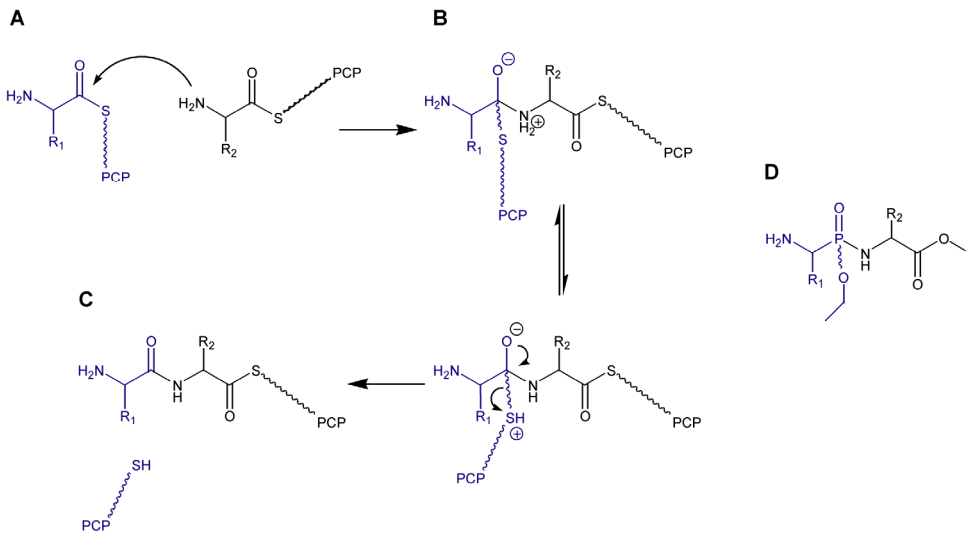


Figure 5.24: Different stages during the amide-bond-forming reaction catalyzed by the C domain. **(A)** Prior to the nucleophilic attack, both acceptor (black) and donor substrate (blue) have sp^2-configured C-terminal carbon atoms. In the transition state, however, the donor's C-terminus has sp^3-type tetrahedral geometry, whose absolute configuration is so far unknown **(B)**. After the condensation reaction, the planar (sp^2) amide bond is found in the product. **(D)** The stable transition state analog suggested in this work.

Additionally, the sulphur atoms were altered to oxygen atoms to decrease potential leaving group qualities and/or stabilize the resulting phosphonic and carboxyl ester functions, respectively. Furthermore, some simplifications were made: The

phosphopantetheinyl residues were replaced by ethoxy- or methoxy groups, and the side chains of both the donor and acceptor amino acid analogs were reduced to mere methyl groups to allow for more generality while maintaining the stereoinformation (figure 5.24D). In total, the transition state analog carries three stereocenters, and therefore 8 diastereomers of the molecule are possible. Two of these stereocenters are pre-determined by the natural donor and acceptor substrates while the preferred absolute configuration of the one in the middle is unknown.

The synthetic strategy for the transition state analog is similar to a synthesis previously published [Sellergren 2000]. It starts with the commercially available and enantiomerically pure *R*- or *S*-1-amino-ethylphosphonic acid, whose amino group was protected with tert-butyl-oxycarbonyl (BOC) in the first step (figure 5.25).

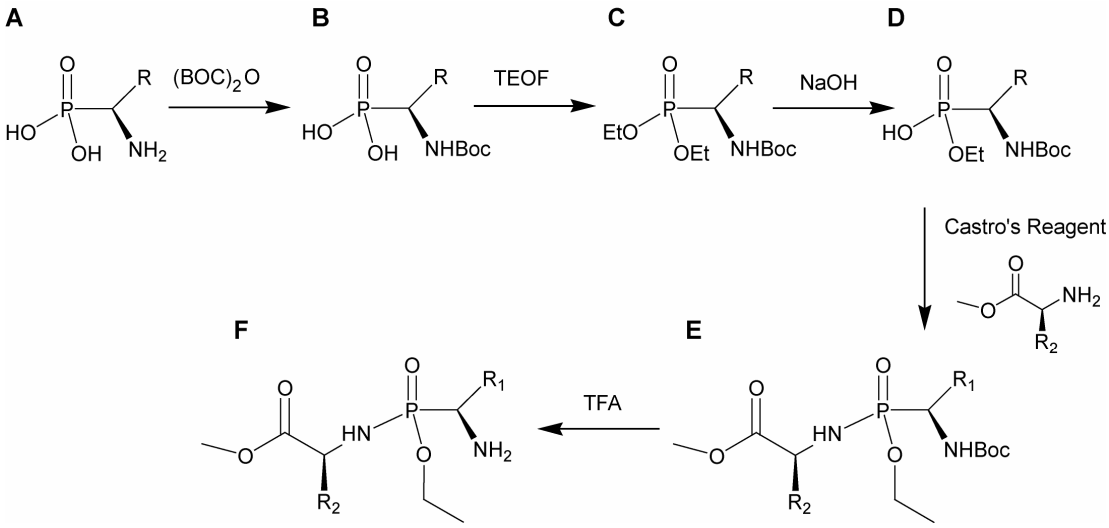


Figure 5.25: Scheme of the synthetic strategy used to obtain the transition state analog. **(A)** The α-amino group is protected with BOC, before the phosphonic acid **(B)** is esterified using TEOF. The resulting diethylester **(C)** is then partially hydrolyzed with NaOH, giving rise to the monoester **(D),** which is consequently coupled to the acceptor aminoacylester with a PyBOP homolog. **(E)** The coupled compound leads to the desired transition state analog **(F)** by acidic cleavage of the protective group.

The N-protected phosphonic acid was then converted into its corresponding diethylester by heating in triethylotho-formiate (TEOF). In the third step, a partial hydrolysis of the diethylester was carried out under basic conditions (NaOH) in ethanol. During this step, an additional stereocenter (phosphorous atom) is unselectively formed so that the reaction product is a mixture of two diastereomers. The resulting hydroxyl group was activated by Castro's reagent (*i.e.* a PyBOP variant) and coupled to either *S*- or *R*-Ala-OEt –

introducing the third stereocenter. Deprotection with TFA in the last step led to the desired products (as determined by ^{1}H-, ^{13}C-, and ^{32}P-NMR) in yields ranging from 10 to 14 %.

In case of proper recognition by – or affinity to – the catalytic site of the C domain, such a stable transition state analog might be able to function as a competitive inhibitor in the condensation reaction and as a suitable substance for co-crystallization experiments.

5.2.9 Inhibitory Assays using TycA and TycB1

The first two modules of the tyrocidine NRPS, TycA and TycB1, appeared to be a suitable test system for inhibitory studies with the synthesized transition state analog. Firstly, their substrates are amino acids (as opposed to a peptide and an amino acid), which means there is a good similarity to the putative inhibitor in terms of size. Secondly, both enzymes interact *in trans*. With respect to the design and proposed function of the inhibitory molecule, the non-covalent nature of competitive inhibitions seems more suitable in the context of two enzymatic partners which also interact non-covalently. Thirdly, the nonribosomal condensation products of two amino acids are known to undergo an uncatalyzed consecutive reaction which leads to cyclization and release from the enzymatic template (figure 5.26) and can be used to monitor the reaction.

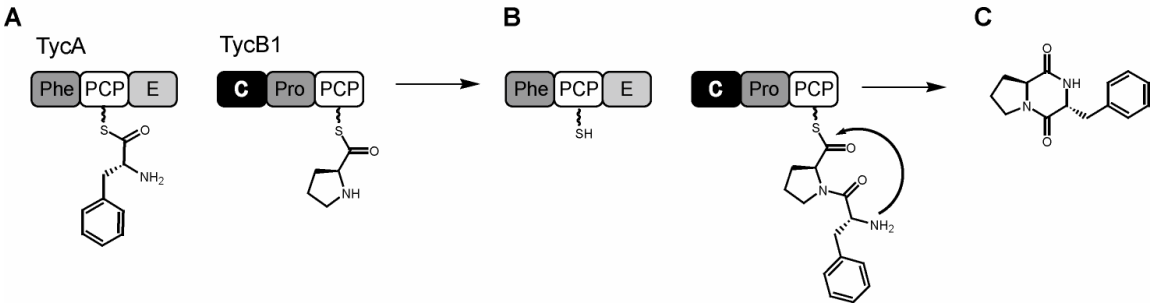


Figure 5.26: Reaction scheme of the DKP producing reaction with TycA and TycB1. The C domain catalyzed condensation reaction between the primed enzymes **(A)** leads to the DPhe-Pro dipeptidyl group bound to PCP_{TycB1} **(B)**. The uncatalyzed intramolecular attack of the intermediate's N-terminus onto the thioester group leads to the formation of diketopiperazine (DKP) species **(C)**.

The resulting molecule has a 2,5-diketopiperazine (DKP) scaffold which is generated by a nucleophilic attack of the N-terminal amino group onto the C-terminal thioester functionality [Bergendahl 2002]. When no further nonribosomal elongation reaction occurs to intercept this side reaction, DKP becomes the main product and can be easily detected from the solution. In the test system used here, the DKP of the dipeptide DPhe-Pro is formed.

Since previous DKP-formation studies had shown that a 2-fold excess of the donor enzyme TycA leads to a faster production of the cyclic dipeptide, 80 µM of *holo*-TycA and 40 µM TycB1 were used for the following experiments. The enzymes were pre-incubated with the *S,S*-configured transition state analog (1 mM) for 15 minutes (25 °C), as this absolute configuration is equivalent to the naturally occurring product intermediate: DPhe-

LPro. Subsequently, both LPhe and LPro were added (200 µM each) along with $MgCl_2$ (10 mM), and the reaction was started by addition of 10 mM ATP. Time dependent samples were taken from the reaction mixture after 1, 5, 10, 20, 60, and 120 minutes and the reactions were stopped by adding TFA – yet no difference in produced DKP amounts could be detected by HPLC-MS analysis; the following figure 5.27 shows two exemplary traces obtained after 20 minutes:

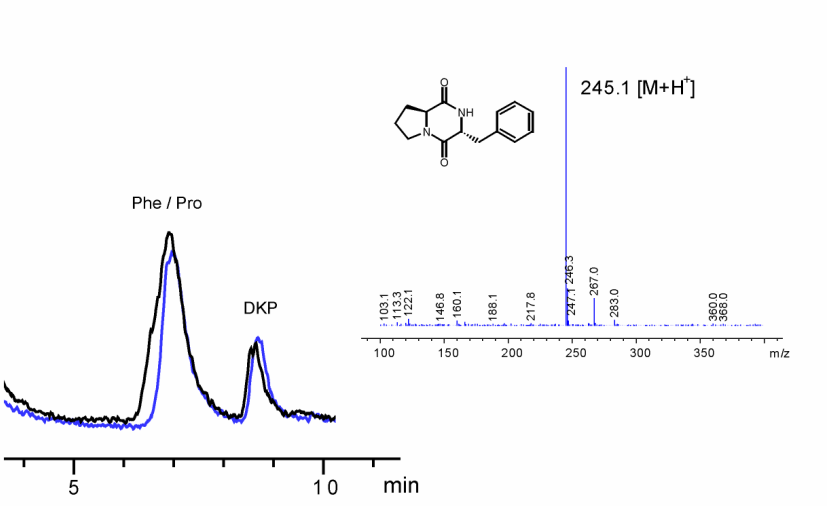


Figure 5.27: Sample LCMS chromatograms (UV) of DKP inhibitory assays with TycA and TycB1 after 20 minutes reaction time. The blue trace shows the reaction outcome in the presence of the synthesized inhibitor, and the black trace is the control without inhibitor. The DKP mass analysis is shown.

5.2.10 Inhibitory Assays using PCP_{TycC5} and C_{TycC6} *in trans*

The cyclization reaction described in 5.2.4 was also tested for inhibition with the synthetic transition state analog. Here, the *R,S*-configured type was used as it matches the natural substrates recognized by C_{TycC6} (LOrn and LLeu). Under otherwise identical conditions (see 5.2.4), the hexapeptidyl substrate DPre-Pro-Phe-DPhe-Asn-Gln was enzymatically transferred onto the PCP domain after 15 minutes of incubation with 1 mM of the transition state analog. However, no effects on product formation could be found. A simple calculation for the local concentration of the hexapeptide within a 20 Å sphere (assuming free rotation of the phosphopantetheine cofactor and neglecting any displacement volume of the enzyme) explains why no effect was seen. The calculation led to a local concentration of ~50 mM – which was to be challenged by 1 mM non-covalent inhibitor. Unless the synthesized transition state analog has extraordinary affinity to the active site, it is constantly displaced by the reactant which undergoes the irreversible

cyclization reaction. Therefore, an *in-trans*-setup of the experiment was chosen, in which the isolated PCP_{TycC5} and C_{TycC6} were used.

Hexapeptidyl-*holo*-PCP_{TycC5} (100 µM) was enzymatically generated *in situ* from the *apo*-enzyme, hexapeptidyl-CoA (125 µM), and Sfp (0.5 µM) by incubation at 25 °C at pH 7.0 for 20 minutes. In parallel, C_{TycC6} was incubated with the transition state analog (*R,S*-configuration). The concentration of the potential inhibitory molecules was adjusted to 100 µM, before the hexapeptidyl-*holo*-PCP_{TycC5} and C_{TycC6} were mixed (60 µM : 1 µM) to initiate the reaction. Time-dependent samples were taken from the reaction mixture after 1, 5, 10, 20, 60, and 120 minutes, stopped, and analyzed by HPLC-MS. In additional (otherwise identical) experiments, several other small tetrahedral or reactant-resembling molecules were used instead of the synthesised phosphonamide to further screen for potential inhibition. The 20- and 120 minute traces of these experiments are shown in the following figure 5.28:

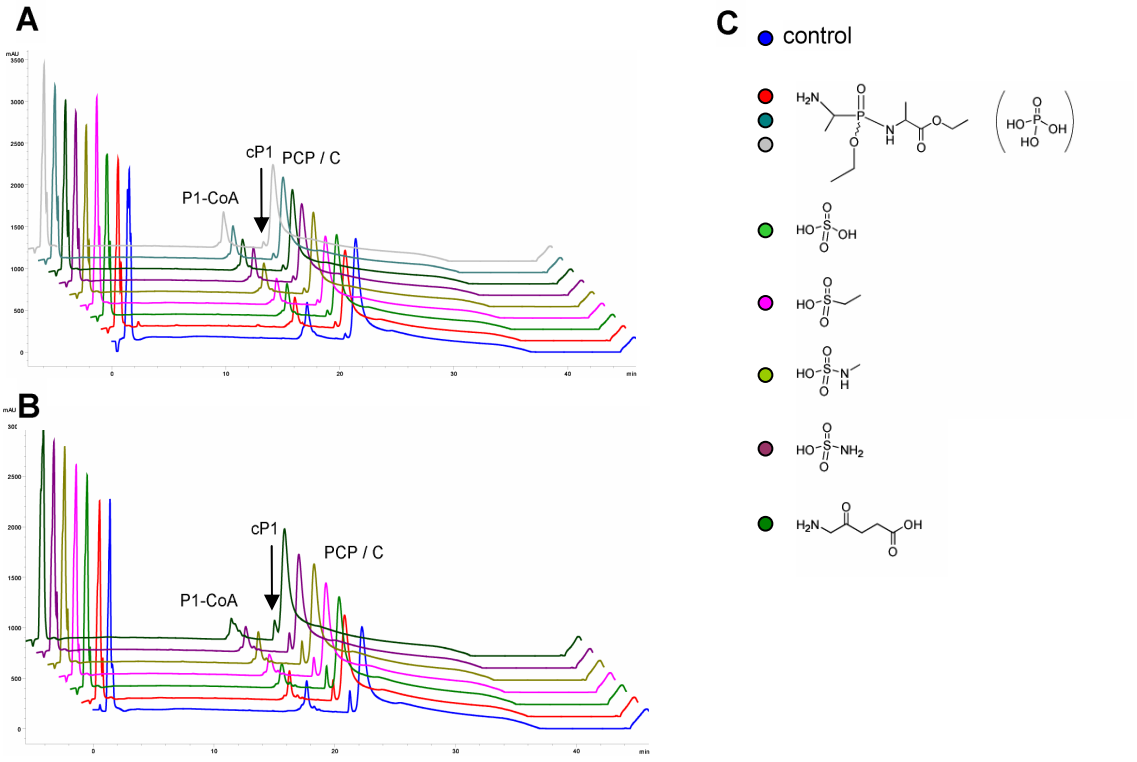


Figure 5.28: Outcome of inhibitory studies using PCP_{TycC5} and C_{TycC6} and various potential inhibitory molecules as shown in UV chromatogram overlays. The signal for the cyclic hexapeptide cP1 after 20 **(A)** and 120 minutes **(B)** is indicated by black arrows. **(C)** The molecules tested are linked to the chromatograms by color-coding. In **(A)**, two additional conditions were tested, in which the final concentration of the (*R,S*)-phosphonamide was raised to 1 mM (grey) and the (*S,S*)-phosphonamide was used (greenblue).

In none of these cases, a measurable difference in product formation was found compared to the control reaction.

6. Discussion

In this work, two different types of enzymes that catalyze amide bond formation in NRPSs have been investigated. Firstly, the formylation (F) domain from the Lgr-cluster of *Bacillus brevis* was addressed. Initially, the F domain had merely been postulated to be responsible for the N_{α}-formylation of the N-terminal valine in linear gramicidin. With the help of enzymatic assays, this hypothesis could be verified. Thus, both the donor and acceptor substrates were identified, and the according specificities were elucidated by substrate alterations. These intermediary results have then been implemented in further experiments aiming at the question, whether a formylated starter unit was necessary for the initiation of the biosynthesis.

Secondly, a better understanding of the condensation (C) domain was sought by means of biochemical characterization and structural studies. So far, the mechanism for the amide/peptide bond forming reactions catalyzed by C domains has not been fully and coherently explained on the molecular level by previously suggested models [Roche 2003, Keating 2002, Bergendahl 2002]. As a suitable test system, the nonribosomal tyrocidine cluster of *Bacillus brevis* was chosen, since it can be considered a "standard NRPS" due to the fact that it works strictly linear and does not depend on external modifying enzymes whatsoever. During earlier studies [Schoenafinger 2003], the bidomain PCP-$C_{TycC5\text{-}6}$ had been found to catalyze an unexpected hexapeptide cyclization reaction. Consequently, this finding served as a starting point for further biochemical experiments, in which a set of other peptides was tested for cyclization at first. One model which states that a conserved histidine residue in the enzyme is necessary for C domain activity [Bergendahl 2002, Roche 2003] was scrutinized and verified in this context by mutational studies. Aside from enzymatic assays, deeper insights into the C domain's nature were sought by crystallographic methods: The PCP-$C_{TycC5\text{-}6}$ bidomain was successfully crystallized and its structure was solved[‡] at a resolution of 1.8 Å. Thus, for the first time, the interactions and the arrangement between two nonribosomal domains could be seen. Besides the reactions performed by single nonribosomal domains, it is these interactions between the catalytic

[‡] Stefan A. Samel and Lars-Oliver Essen, Department of Biochemistry, Philipps-University Marburg, Germany.

entities that will help us understand the mechanics of the nonribosomal machinery and eventually provide us with the greater picture.

In the following subsections, first the F domain and then the C domain are discussed.

6.1 The F Domain

6.1.1 The F Domain in the World of Formyltransferases

Nature's most prominent formylation reaction is carried out by methionyl-tRNA-formyl transferases (FMTs, ~35 kDa) which catalyze the transfer of a formyl group from N^{10}-formyl-tetrahydrofolate onto the α-amino group of methionine bound to the initiator $tRNA^{fMet}$ in prokaryotes [Schmitt 1998]. Consequently, the bacterial proteome is N-terminally formylated. FMTs comprise two subdomains: The N-terminal (~23 kDa) is responsible for the formyltransferase activity, while the C-terminal (~12 kDa) is necessary for the recognition and positioning of the acceptor substrate methionyl-$tRNA^{fMet}$ (Figure 6.1).

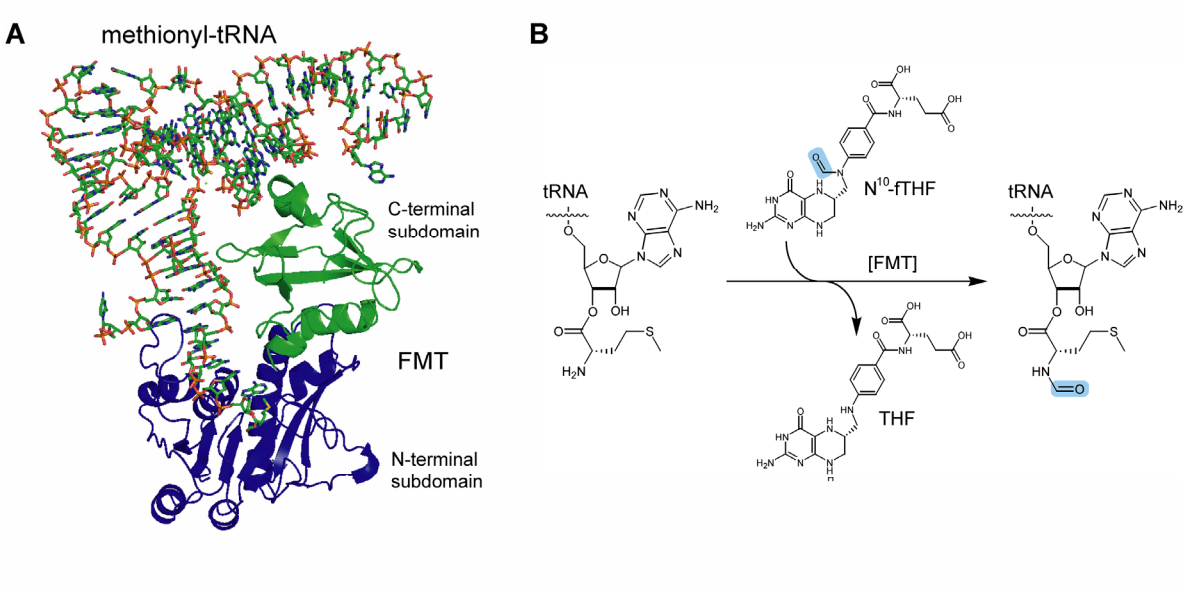


Figure 6.1: (A) Binary complex of FMT and methionyl-$tRNA^{fMet}$ from *E. coli*. The C-terminal subdomain (green) positions the tRNA, while the N-terminal subdomain (blue) has formyltransferase activity. **(B)** Scheme of the reaction catalyzed by FMTs. The formyl group (light blue) is enzymatically transferred onto the α-amino group of the methionyl moiety.

The nonribosomal secondary peptide products, however, are much more versatile and do not generally carry such a modification on the first building block. In fact, formylated α-amino groups are rather exceptional in nonribosomal products and thus far only known in two cases: linear gramicidin and the anabaenopeptilides [Rouhiainen 2000]. When looking at their biosynthetic gene clusters, additional coding regions (600 bps for LgrA, 1400 bps for AdpA) can be found at the 5'-end of the respective first NRPS gene. The resulting N-terminal 23 kDa domain of LgrA exhibits striking sequence similarities to the N-terminal

subdomains of FMTs [Schracke 2005] whereas in AdpA the C-terminal moiety (~270 aa) of that protein sequence appears to be part of a poorly conserved C domain. The N-terminal stretch of 200 amino acids, however, does also exhibit high sequence similarities to FMTs (see appendix).

One common motif can be found throughout N-formyltransferases: A region of fairly well-conserved residues around a characteristic "SLLP" sequence (115 to 160, roughly, in figure 6.2).

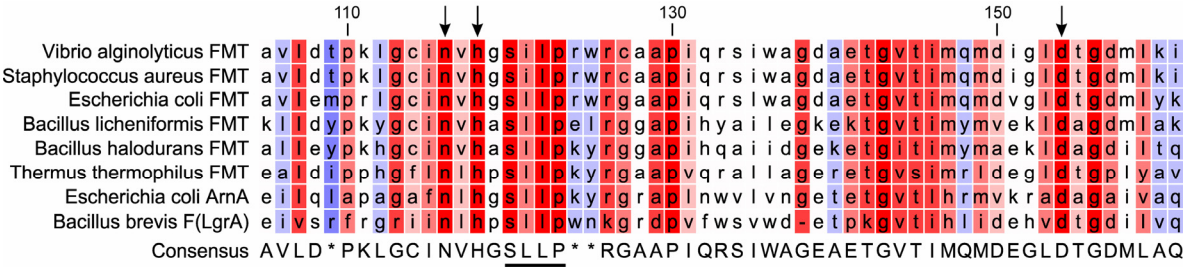


Figure 6.2: F_{LgrA} aligned with several FMTs and ArnA. Red indicates a high, white a medium and blue a low degree of conservation. The "SLLP" sequence is underlined. Residues involved in the formyltransferase catalysis are indicated by arrows: N116, H118, and D154.

In 2005, the bifunctional enzyme ArnA from *E. coli* was co-crystallized with substrate analogs [Williams 2005]. The N-terminal formyltransferase moiety of ArnA catalyzes the transfer of a formyl group onto the amino group of UDP-bound 4-amino-4-deoxy-L-arabinose by consumption of N^{10}-formyl-tetrahydrofolate. Since the natural formyl donor substrate was unstable, the N^5-formylated variant was used in the crystallization experiments along with UDP monophosphate. With the 1.2 Å structure presented in figure 6.3, several residues could be identified in the active site that were previously suggested [Morikis 2001, Shim 1998] to be critical for formyltransferase activity: N116, H118, and D154 (Figure 6.3), all of which are strictly conserved throughout formyltransferases.

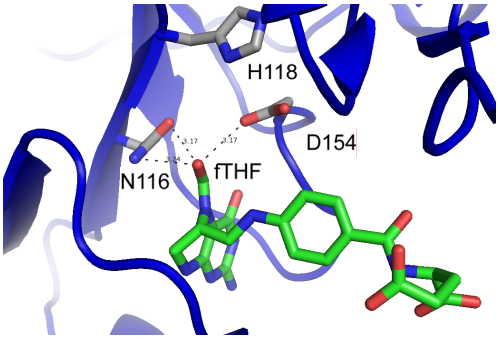


Figure 6.3: N^5-fTHF bound to the acive site of ArnA. The conserved residues N116, H118, and D154 are shown as stick models. N116 and D154 are in close proximity to the formyl group oxygen atom.

Thus, a plausible model for the catalytic mechanism of formyltransferase activity can be proposed (figure 6.4): Due to the fact that a mutation of Asn116 to Asp resulted in a complete loss of activity [Shim 1998], its location proximal to the carbonyl oxygen of the formyl group suggests a critical role in charge stabilization of the oxyanion of the transition state after the nucleophilic attack of the amino group. His118, however, is believed to capture the surplus proton of the acceptor amino group during this step, and it is temporarily transferred to Asp154 before it is passed on to N^{10}/N^5 of the donor substrate at the end of the reaction.

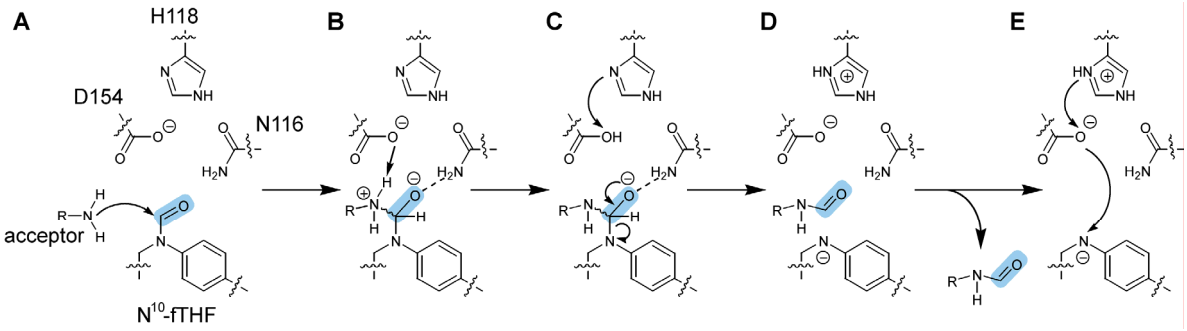


Figure 6.4: Scheme of the proposed catalytic mechanism of the ArnA formyltransferase. **(A)** The initial attack of the acceptor nucleophile. **(B)** The tetrahedral transition state is stabilized by N116 and the excess proton transferred to D154. **(C)** The proton is translocated to H118, and the elimination of the product is initiated. **(D)** The formylated amino species can leave the active site, and **(E)** the surplus proton is shuttled from H118 via D154 to the mesomerically stabilized tetrahydrofolate anion.

Since the natural substrate N^{10}-fTHF could not be used, it remains unclear whether there are significant differences in the natural setting. The relative positioning of both N^5 and N^{10} as seen in the crystal structure, however, does not suggest any dramatic change of the location of the formyl group. In fact, it appears that the position of the formyl group itself hardly varies between the N^5- and N^{10}-fTHF. In this work, both cofactors could be used to formylate PCP-bound valine with F-A-PCP_{LgrA1}, even though the reaction was 18-fold faster when the N^{10}-type cofactor was provided. Overall, the similarities in sequence, size and the preference for N^{10}-fTHF as donor substrate, suggest that the newly described nonribosomal F domain (F_{LgrA}) is structurally and functionally homologous to its counterparts from the primary metabolism. One might therefore speculate that it had been adopted rather than separately evolved to serve in the highly specialized context of NRPSs.

6.1.2 Acceptor Substrate Specificity of the Nonribosomal F domain F_{LgrA}

To investigate the substrate tolerance of F_{LgrA}, a variety of amino acids, both cognate and non-cognate to the A domain in LgrA1, were enzymatically loaded onto *apo*-F-A-PCP_{LgrA1} for formylation experiments. It was found that in accordance with the A domain's specificity the three branched aliphatic amino acids valine, leucine, and isoleucine were formylated. A relative comparison of the product amounts formed after a defined time clearly showed that valine (100%) was preferred over leucine and isoleucine (~5% both). This finding is empirically supported by the natural distribution of variants within the mixture of linear gramicidins (gramicidin D) extracted from *Bacillus brevis* cultures [Hotchkiss 1940], even though leucine is not mentioned as a substitution for valine in position 1. At this point, it cannot be judged whether the fact that leucine was accepted is an artifact of the experimental setup (excised enzyme, *in vitro* conditions) or the analysis of gramicidin D itself was error prone or incomplete. To this end, it is worth mentioning that both Ile and Leu would give identical mass signals, similar NMR signals, and possibly retention times in an HPLC chromatogram that are difficult to distinguish.

As for the non-cognate amino acids, none of the potential substrates were N-formylated even though a broad variety in terms of physical and chemical properties was probed (Figure 6.5). As a conclusion, the A domain's specificities appear to be reflected upon the F domain's. This does not have to be by design – instead it might be a simple necessity of the acceptor substrate binding site of the F domain in order to retain its affinity in the context of spatial limitations.

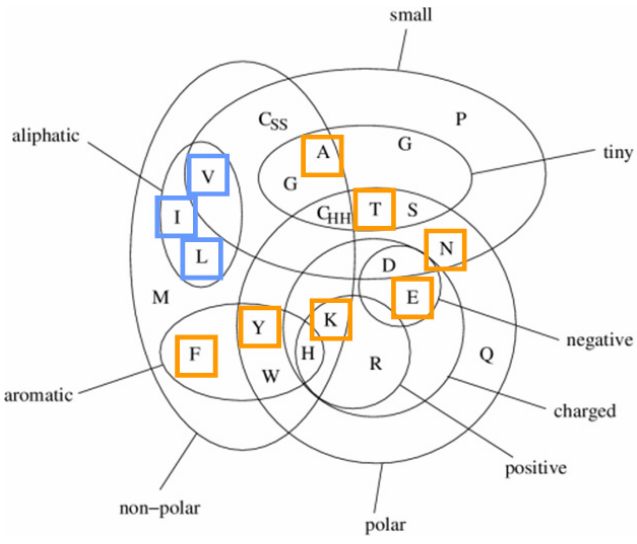


Figure 6.5: Amino acyl substrates tested for formylation with F-A-PCP_{LgrA1}. Only the aliphatic branched amino acids valine, leucine, and isoleucine were modified (blue) while seven other substrates (orange) were not accepted by the F domain.

6.1.3 The Necessity of Formylation in Linear Gramicidin

Since linear gramicidin is always found N_{α}-formylated, it was tested if the reaction catalyzed by the F domain is essential for the initiation of the complete nonribosomal biosynthesis. The hexadomain enzyme F-A-PCP-C-A-$PCP_{LgrA1-2}$ was therefore constructed and used in an assay where the formylation and condensation reaction are competing. First, the *holo*-enzyme was supplied with its cognate substrates valine, glycine, and ATP. After the A-domain-catalyzed activation and thiolation of Val1 and Gly2, respectively, the condensation reaction would then theoretically be able to occur without prior formylation of Val1. However, no dipeptide was generated without the formyl donor, and only when the cofactor was present, formylated dipeptide was produced. This could be interpreted in two ways: Firstly, the F domain might exhibit a substantially higher affinity for the acceptor substrate (PCP-bound valine) than the donor site of the C domain. Secondly, the C domain might be highly specific for formylated donor substrates as opposed to non-modified ones. The outcome itself is of great value for the producing organism, as it can be expected that the production of non-formylated linear gramicidin would lead to biologically inactive peptides at a great metabolic cost. Mainly two reasons for the importance of the N-terminal modification of each monomer can be envisioned: Without formylation, the N-termini would be cationic ammonium groups under physiological conditions. Therefore, an electrostatic repulsion between the two monomers could occur – impeding the formation of the dimer. Furthermore, the presence of formally charged groups is contrary to the concept that hydrophobicity is needed for the secondary metabolite to find its way into its biological target, lipid bilayer membranes. On the other hand, the stability of the dimeric complex is greatly increased by two additional hydrogen bonds at the interface when the amino acids at postion 1 are formylated (Figure 6.6).

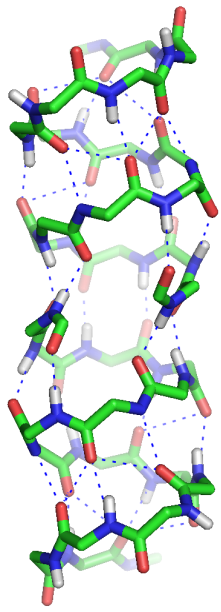


Figure 6.6: The backbone of dimeric linear gramicidin. The formyl groups at the dimer interface contribute to the network of hydrogen bonds that stabilizes the complex.

6.2 The C Domain

The condensation domain fulfills a central role in NRPSs, as this enzymatic unit is generally required for amide/peptide bond formation during the assembly of nonribosomal secondary metabolites. The starting point for this work here was the bidomain enzyme PCP-$C_{TycC5\text{-}6}$ from the tyrocidine NRPS which had been found to cyclize the hexapeptide DPhe-Pro-Phe-DPhe-Asn-Gln earlier [Schoenafinger 2003]. This observation was unexpected as the assay had initially been aimed at an elongation reaction through the presence of the C domain's natural acceptor substrate leucine – presented in various forms *in trans*. Even when the concentrations of such potential acceptor substrates (leucine, Leu-SNAc, Ala-SNAc, DAla-SNAc, glycine, ethylene-diamine, and even leu-*holo*-PCP_{TycC6}) had been increased to 20-fold excess over PCP-$C_{TycC5\text{-}6}$, no elongation was observed. Since the only suitable nucleophile in the donor substrate is the α-amino group of DPhe1, it was hypothesized that the cyclic hexapeptide exhibits a head-to-tail connectivity. This was verified in two ways: firstly, by MS^n spectrometric analysis, in which a fragment of four amino acids was found that originates from Gln-Phe-Pro-DPhe (calc.: 503.2294, obs.: 503.2299); and secondly, by using an N-terminally acetylated variant of the hexapeptidyl substrate which led to no product formation in the assay.

According to the common understanding of C domains so far, they have separate binding sites for the donor and acceptor substrates: The donor substrate is provided from the upstream and the acceptor from the downstream PCP domain – a situation which is not given in the cyclization experiment with hexapeptidyl-*holo*-PCP-$C_{TycC5\text{-}6}$. One possibility for the observed cyclization was that it may not be actively catalyzed by the C domain. Instead, it could be a side effect of either the loading onto the PCP domain or the altered environment altogether. To test this, the hexapeptide was loaded onto the isolated PCP_{TycC6}, which led to no formation of a cyclic or even hydrolytic product under otherwise identical conditions. A second test was performed by using a mutant of PCP-$C_{TycC5\text{-}6}$, in which the conserved, supposedly catalytic histidine residue (H224) [Bergendahl 2002, Roche 2003] had been exchanged by alanine in one case and valine in the other. Both resulting proteins were well-soluble and their α-helical CD spectra were in good

agreement with the wildtype enzyme. When these mutants were used in analogous cyclization experiments, no cyclized hexapeptide was produced.

Further investigations with variations of the DPhe-Pro-Phe-DPhe-Asn-Gln substrate revealed that Gln6 could not be replaced by alanine at the C-terminus and a D-configured N-terminal amino acid is necessary for cyclization to take place. When the chain length was increased to 8 by insertion of two alanines after position 4, the linear substrate was only released by C-terminal hydrolysis. Unfortunately, substrates with a C-terminal ornithine (as in the natural substrate) were not accessible due to instant cyclization of the peptide moiety in the corresponding CoA species, likely mediated by the intramolecular attack of the free δ-amino group onto the thioester, which leads to a six-membered lactam. However, glutamine appeared similar enough, as its only difference to ornithine is the carbonyl oxygen at the γ-position which abolishes most of the nucleophilicity of the δ-amino group since the latter is now part of a mesomerically stabilized amide with a partial positive charge on the nitrogen atom. The fact that DPhe1 could not be replaced by LPhe or LAla, on the other hand, was more puzzling, because the natural acceptor substrate leucine is L-configured. For the cyclization of a hexapeptide in a head-to-tail manner, a pre-folding of the linear precursor must be possible to allow for a proximal orientation of both termini. It can be assumed, that the delicate sequence of L- and D-configured residues – especially, when sterically demanding such as phenylalanine – can have a dramatic effect on such a pre-folding. For the tyrocidine linear decapeptide a β-sheet like folding has been reported [Bu 2002], and following the "4n+2-rule", an analogous secondary structure can easily be envisioned for the hexapeptide, too. Thus, obviously both the C domain's catalytic residue H224 and an intrinsic effect of the substrate itself contribute to the observed phenomenon.

These findings led to a deeper scrutiny of the previously suggested catalytic models of the C domain. The first one states that the second histidine of the core 3 motif HHxxxDG (this will be referred to as H224 – like in PCP-$C_{TycC5-6}$, to avoid confusion) acts as a base catalyst required to increase nucleophilicity of the acceptor α-amino group by either capturing the excess proton after the initial attack or the abstraction of a proton from the corresponding α-ammonium group before the attack onto the thioester group.

For C_{TycB1} it was experimentally shown by mutational studies that H224 is essential for catalytic activity [Bergendahl 2002], and analogous findings were made here in the context of the cyclization reaction with PCP-$C_{TycC5-6}$. However, Keating and coworkers made contrary observations when testing the stand-alone C domain VibH [Keating 2002]. This C domain condenses PCP-bound dihydroxy-benzoate with the primary amino group of free norspermidine as part of the vibriobactin nonribosomal biosynthetic process in *Vibrio cholerae*. When H224 was mutated to alanine, merely a 90% decrease in activity was observed – indicative of a non-essential role of the respective residue. It is argued that solely the positioning of both reactants by the C domain leads to the formation of the thermodynamically more stable amide.

In the bidomain structure of PCP-$C_{TycC5-6}$, a buffer-derived sulfate ion is seen in proximity to H224 ($N_{\varepsilon 2}$, 2.87 Å). The pK calculation for the side chain of H224 (pK 11.8) indicates that it is protonated in this environment (H++ server, [Bashford 1990, Gordon 2005]). Thus, it cannot act as Brønsted-base to perform catalysis. Instead, the binding of sulfate in this position suggests an electrostatic interaction between the negatively charged sulfate and the imidazolium group. Taking a closer look at the active site, two additional features are found that stabilize the sulfate ion in its very position (Figure 6.8): A hydrogen bond with the amide nitrogen of G229 and the dipole moment of helix 4.

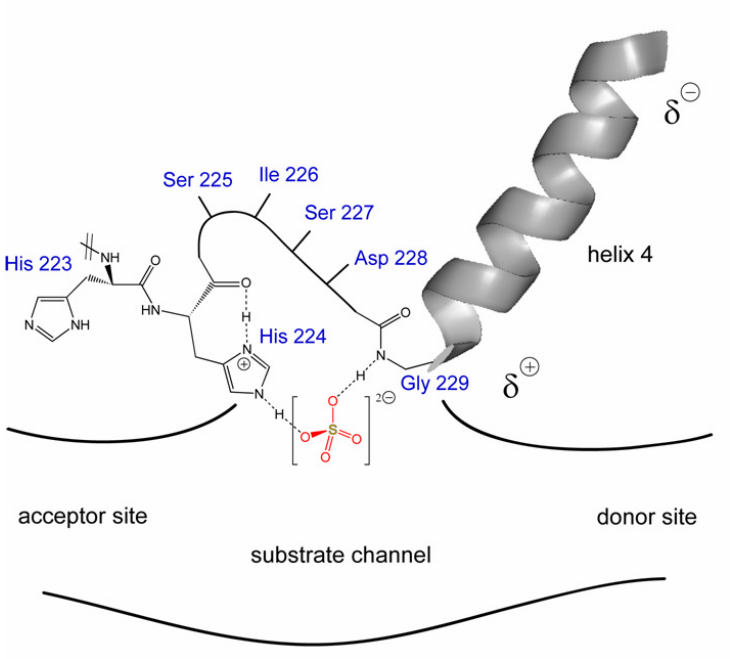


Figure 6.8: The chemical environment at the C domain's active site as seen in the bidomain structure. A sulfate ion is bound to H224 and G229 by polar interaction/hydrogen bonding. The dipole moment of helix 4 further stabilizes the sulfate. H223 is turned away from the substrate (solvent) channel.

The first histidine in the core 3 motiv HHxxxDG is found to be embedded in a hydrophobic pocket that is turned away from the substrate channel, making an involvement in the catalytic process rather unlikely, unless substantial structural re-arrangements take place. Accordingly, its calculated pK value of -3 is very low.

With these findings, a different catalytic model for the C domain can be suggested here if one assumption is made: The negatively charged, tetrahedral sulfate ion found in the crystal is analogous to the transition state of the amide bond forming reaction in the C domain. This transition state – generated by the nucleophilic attack of the α-amino group on the thioester carbon atom – is also sp^3-configured and carries a negative charge (oxyanion). The structural features and interactions that can be held accountable for the binding of sulfate in the crystal can in turn be pictured as the enzymatic environment that stabilizes the transition state. As a result, the activation energy is lowered and the reaction speed increased. This model does not contradict the previous findings that H224 is essential for TycB1 and TycC6 while it is not for VibH. The fact that *three* interactions with the enzyme separately contribute to the transition state stabilization leaves room for the possibility that in individual cases only *two* of them might be sufficient.

One reason for the postulate of a Brønsted-base in the catalytic site of the C domain originates from the concept that primary amines are expected to be practially quantitatively protonated under physiological conditions – and therefore cannot act as nucleophiles. When taking a closer look at several pK values, however, it becomes evident that this argument can be refuted. While aliphatic primary amines have pK values between 9.2 and 10.6 [Hall 1957], the α-amino groups of oligopeptides are less basic ($pK_{Gly\text{-}Gly}$ = 8.31, $pK_{Ala\text{-}Ala\text{-}Ala}$ = 8.03), and although values for aminoacyl thioesters could not be found in the literature, the pK value of glycine oxoester is reported to be 7.75 [Hall 1957]. Still, the majority of amino groups are protonated, but the un-protonated portion is able to perform the condensation reaction without a base, and the fact that the reactant is withdrawn from the equilibrium drives the reaction. This is consistent with the known issue of C domains being very slow catalysts compared to A domains, for instance (~1:1000).

The overall architecture of both structurally known C domains, C_{TycC6} and VibH, clearly shows that there is no classical deep cavity for substrate binding but rather a canyon-like

groove in which the substrates are positioned. The natural substrates for VibH (PCP-bound dihydroxy-benzoate and norspermidine) and C_{TycC6} (nonapeptide and leucine, both PCP-bound) vary greatly in size and shape, yet the enzymatic environments for their positioning are analogous. It is therefore expected that the Ppan cofactor coordination with the enzyme along the solvent channel is critical for the positioning of the reactants. Unfortunately, these interactions are not seen in the *apo*-PCP-$C_{TycC5-6}$, but a structural superposition of the carnitine acetyltransferase/CoA complex [Hsiao 2004] with the C domain's N-terminal subdomain places the pantetheine moiety along the floor of the canyon in such a way that the thiol group almost coincides with the sulfate ion at the active site (Figure 6.9).

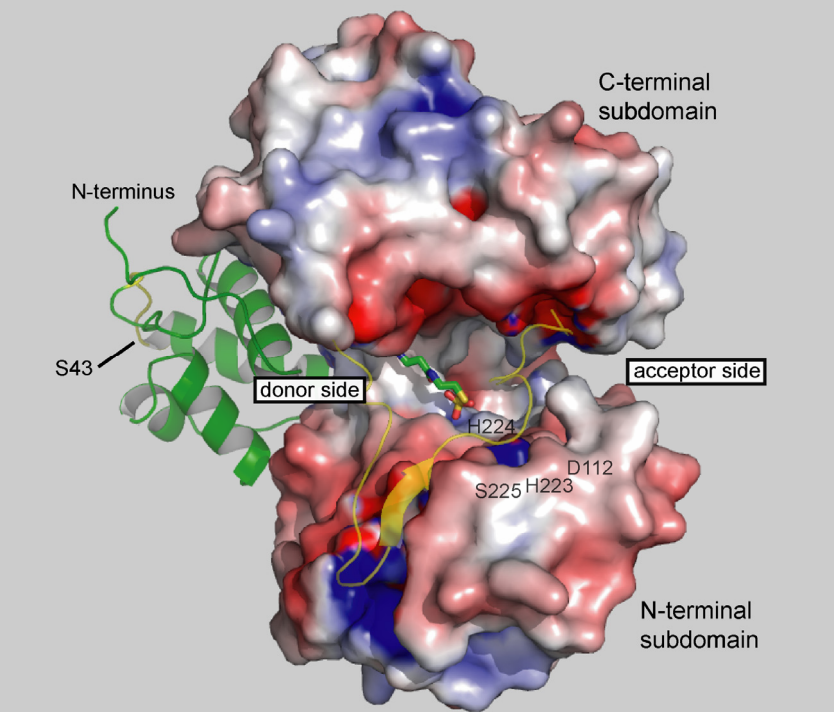


Figure 6.9: The bidomain structure of PCP-$C_{TycC5-6}$ with Ppan modeled into the solvent channel. The PCP domain is shown as cartoon (green), and the C domain's surface plot is colored by its charge (red = acidic, blue = basic). The bridging region has been reduced to a transparent cartoon (yellow) for clarity. The Ppan arm thiol group almost coincides with the sulfate ion bound to H224.

In the inhibitory studies perfomed with phosphonamide-based transition state analogs, such a positioning is not possible, as the molecules used do not carry a phosphopantetheine group. Consistently, inhibition was not observed when the cyclization experiment with the hexapeptidyl-*holo*-PCP-$C_{TycC5-6}$ was made. The problems of a high local concentration of the substrate presented *in cis* (see chapter 5) and the single turnover nature of the reaction were sought to be overcome by using the separated domains. The cyclization reaction in the according *in trans* experiments, however, could also not be inhibited by the transition

state analogs. Likely, larger molecules with substantial portions of the Ppan arm on both the donor and acceptor sides could be more promising candidates for C domain inhibition.

6.2.1 Comparison to the Ribosomal Peptidyl Transferase Site

The C domain's counterpart in the primary metabolism is the peptidyl transferase (PT) site of the ribosome. As in NRPSs, the substrates consist of amino acyl and peptidyl groups bound to a carrier molecule. These carrier molecules are transfer-RNAs (~25 kDa) from the solution which bind their substrates *via* an oxoester bond between the 3'-hydroxyl group of the 3'-terminal ribose moiety of adenosine and the substrate's C-terminus. As for NRPSs, the *holo*-PCP domains (~11 kDa) implement covalent substrate binding by a thioester bond between the Ppan cofactor's sulfhydryl group and the carboxylic group of the substrate. In analogy to the C domain – whose reactants are generally supplied from two directions –, the peptidyl transferase site of the ribosome is provided with its substrates from the neighboring aminoacyl (A) and peptidyl (P) sites. Chemically speaking, both enzymes perform the aminolysis of an ester bond to produce a peptide bond (Figure 6.10).

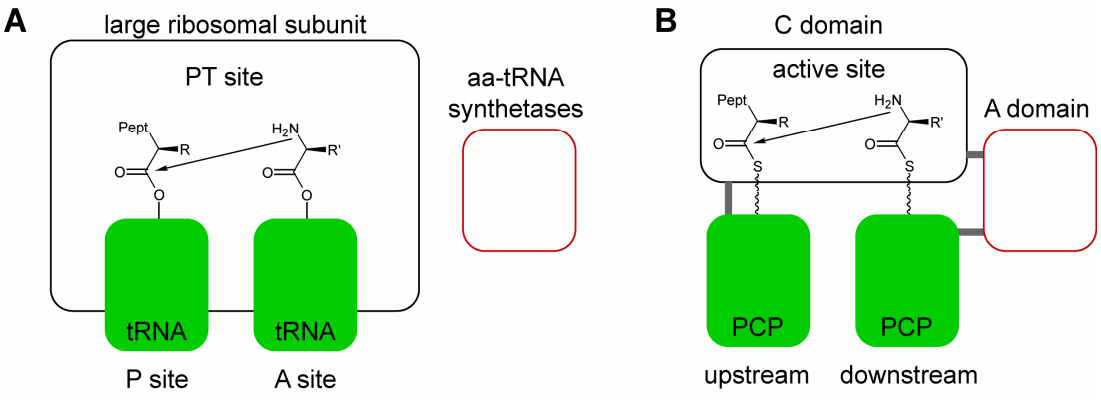


Figure 6.10: Schematic comparison between the ribosomal and nonribosomal peptide bond forming systems. **(A)** The large ribosomal subunit binds its soluble substrates in the A and P sites while the domain sequence in NRPSs determines which reactants are presented to the C domain **(B)**. The functionally homologous carrier molecules are shown in green. The sequence of the nonribosomal domains is indicated by connecting lines (grey). Substrate activating enzymes (aminoacyl-tRNA-synthetases and A domains) are shown in red.

When the 2.4 Å structure of the large ribosomal subunit from *Haloarcula marismortui* was presented in 2000 [Ban 2000], a first view of the PT site had been made possible. Today, we know that the catalytically active site consists of rRNA, rendering the ribosome a ribozyme: There are no peptidic elements within a 15 Å radius of the PT site. Instead, the

nucleotides surrounding the active site are highly conserved (figure 6.11A), and no divalent metal ions such as Mg^{2+} are present [Schmeing 2005]. The PT site has since been subject to extensive investigations [Beringer 2007], involving mechanistic experiments and co-crystallizations with substrate analogs. As a consequence, a model for the reaction has been proposed [Sato 2005] (figure 6.11B). It is believed that the transfer of the donor substrate is initiated by the attack of the acceptor substrate amino group – but in comparison to the addition/elimination reaction suggested for C domains, a concerted proton shuttle mechanism *via* a six-membered transition state appears likely. The strictly conserved ribosomal residues (C2063, A2451, and U2584) do not participate in the chemical catalysis but contribute to the delicate network of hydrogen bonds (which involves two water molecules) required for transition state stabilization.

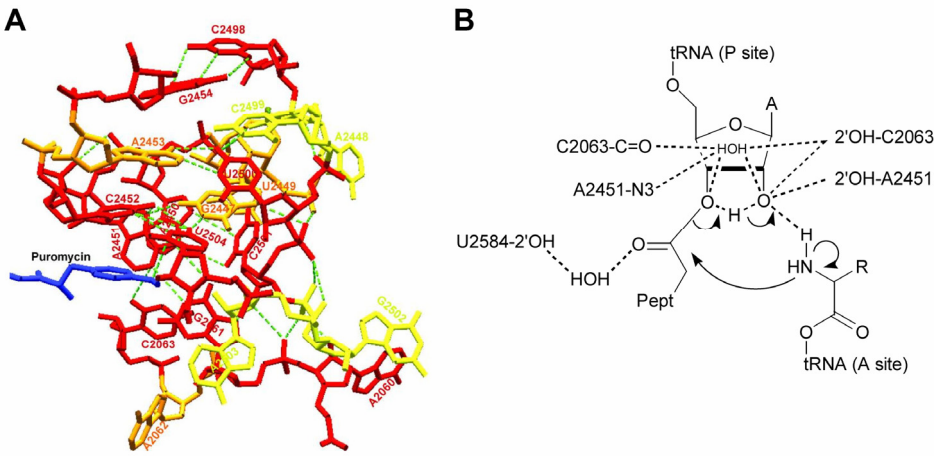


Figure 6.11: (A) Degree of conservation in the PT site. The active center was mapped by co-crystallization with puromycin. Source: [Sato 2006]. Residues shown in red cannot be substituted by any other base while retaining PT activity. Orange residues tolerate an exchange to one other base and yellow ones to two or three. **(B)** The concerted proton shuttle mechanism. A delicate network of hydrogen bonds is realized by three conserved bases and two water molecules.

As in nonribosomal C domains, it is probable that the enzyme merely positions the reactants in an environment favorable for the reaction to occur as opposed to getting covalenty involved in reaction intermediates or transition states. In both cases, obviously no Brønsted-base is necessary.

Even though both the ribosome and NRPSs produce peptides, they are designed to serve different purposes and are thus specialized in their own fields. In this context, it is stunning how analogously the two machineries are functionally constituted: They both accept carrier

molecule bound substrates which remain covalently attached until the termination of the syntheses, and their catalytic entities that perform elongation reactions do not form covalent intermediates with the nascent product chains. In fact, they do not have to since the critical substrate activation steps are performed by other enzymes (aminoacyl-tRNA-transferases and A domains). In both cases, ATP is consumed to form mixed anhydrides before the ester bond is generated with the carrier molecules. In comparison to the amide bonds in the products, these esters appear unstable enough to allow for a condensational catalysis solely driven by the substrate positioning and the electrostatic environment provided by the enzymes.

6.2.2 The Transglutaminase Homolog AdmF

Recently, an outstanding discovery was made that expands the amide bond forming toolbox for secondary metabolite production by another type of enzyme: transglutaminase (TG) homologs. Within the andrimid biosynthetic cluster *adm*, the TG homolog enzyme AdmF (35 kDa) was found to catalyze the condensation reaction between two carrier protein bound substrates *in trans* [Fortin 2007]. Even though the polypeptide-crosslinking *via* a glutamine side chain catalyzed by TGs is well known [Pedersen 1994, Griffin 2002], the occurrence of such a catalyst within the hybrid PKS/NRPS system of *adm* was surprising, since TG-type enzymes had not been recognized as enzymes involved in antibiotic biosyntheses before. AdmF transfers the acyl moiety of its donor substrate octatrienoyl-AdmA to the active cysteine 90 in the first step, forming a covalent thioester intermediate. The octatrienoyl group is subsequently transferred to the amino group of the acceptor molecule β-phenylalanyl-AdmI (Figure 6.12).

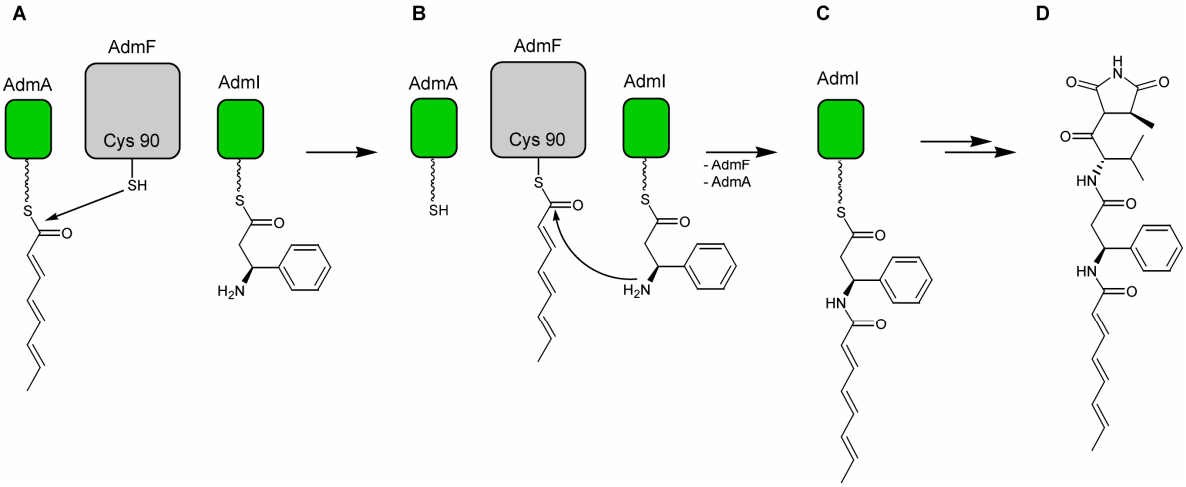


Fig. 6.12: Scheme of the AdmF reaction. **(A)** Transfer of the octatrienoyl-group from AdmA onto Cys 90 of AdmF. **(B)** Aminolysis initiated by the nucleophilic attack of the β-amino group of the AdmI-bound acceptor substrate leads to an amide bond **(C)**. After subsequent steps, the final product andrimid **(D)** is generated.

Cysteine 90 is part of a catalytic triad (Cys-His-Asp), and even though the mechanism is not known for this case, it can be suspected that it is analogous to the aminolysis step in the transglutaminase reaction [Pedersen 1994]: The oxyanion in the tetrahedral transition state is stabilized by hydrogen bonds while the neighboring histidine captures the excess proton.

Functionally, AdmF is homologous to free standing C domains, such as VibH. It remains unclear, why TG-type catalysts were adopted to serve in the context of secondary

metabolite production, yet the fact that AdmF recognizes ACP donors and PCP acceptors draws attention to its specialized capabilities of carrier protein recognition. Moreover, the acceptor substrate β-amino group differs from the α-amino groups usually recognized by C domains. With respect to the typical side chain amino group acceptors in standard TG reactions, the substrate recognition might be more favorable when TG homologs instead of C domains are used.

6.3 Inter-Domain Communication

The interactions between nonribosomal domains are critical for product formation. Even though the reactions catalyzed by the essential set of domains (A, PCP, and C) are fairly well understood, their interactions during several different stages of synthesis are so far unknown. However, a certain hierarchy of the reactions can be postulated: The basic requirement for functionality is the *apo*-to-*holo* conversion of the PCP domains by interaction with a Ppan transferase such as Sfp. On the second level, the building blocks need to be activated by the A domains. Now, the *holo*-PCP domains must be loaded with the proper substrates, before they are translocated to the C domain. It becomes evident, that the carrier proteins need to interact with at least three types of enzymes: the Ppan transferase, the A domain and the C domain. From the second module on, PCP domains even have to interact with one additional C domain – depending on whether its substrate is the donor or acceptor in the reaction (figure 6.13). In nature, not only the essential domains occur, but also modifying domains such as thioesterases, epimerases, N-methyl and formyltransferases, reductases, oxigenases, and the proof-reading thioesterase II, for example.

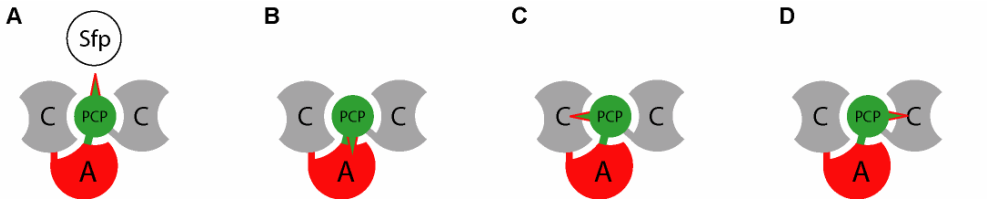
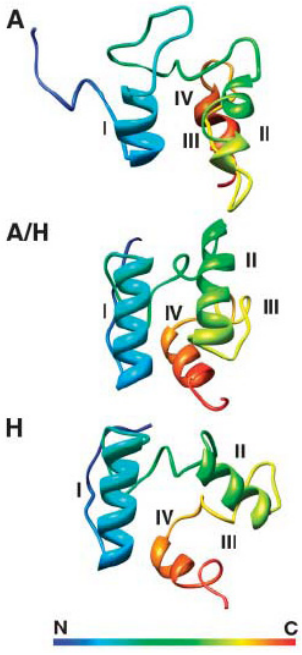


Figure 6.13: Schematic of the interactions between a PCP domain and other enzymatic entities during different stages of nonribosomal peptide synthesis. **(A)** The *apo*-to-*holo* conversion by interaction with the Ppan transferase Sfp. **(B)** The *holo*-enzyme is primed with an A-domain-activated amino acid, which is then **(C)** translocated to the upstream C domain to serve as acceptor. **(D)** After the peptide bond formation the carrier molecule positions the donor substrate at the downstream C domain.

If we assumed an equal affinity of the PCP domain to all of these, the amount of bio-inactive side products would be vast and their production quite costly to the producer. In fact, NRPSs are highly accurate and this can only be achieved by substantial control of the reaction timing.

In order to perform all of its tasks, the PCP domain must furthermore be able to physically reach the active sites of all interaction partners. Even though supplied with the

stretchy ~20 Å Ppan cofactor, this is difficult to achieve unless the relative arrangement of the corresponding domains is somehow situationally optimized. NMR titration experiments [Koglin 2006] with ^{15}N-labeled *apo*- and *holo*-PCP$_{TycC3}$ in the presence of either Sfp or a type II thioesterase led to the discovery of three different folds for the PCP domain (Figure 6.14).



While the A-fold is only seen with the *apo*-PCP upon titration with Sfp, the H-fold appears to be reserved for the *holo*-carrier protein and it is induced by interaction with the thioesterase II. Nevertheless, the both *apo*- and *holo*-forms have a common intermediate folding (A/H), which exhibits the highest degree of secondary (α-helical) structure. This A/H-fold had been previously observed when the excised PCP$_{TycC3}$ structure was determined by NMR from the solution without any interaction partner [Weber 2000]. Furthermore, its overall fold resembles the topology of acyl carrier proteins (ACPs) from *Escherichia coli* fatty acid synthase.

Figure 6.14: The three conformational states of PCP$_{TycC3}$, as observed by titrations with Sfp and type II thioesterase [Koglin 2006].

In the *apo*-PCP-C$_{TycC5-6}$ structure presented in this work, the PCP domain is clearly seen in the A/H-fold. The relative arrangement of the domains indicates that their constructive interaction – where PCP would donate a substrate to the downstream C domain (figure 6.13D) – is not represented in the crystal structure. With a measured distance of 49.6 Å between the PCP domain's active serine 43 and histidine 224, the active site cannot be reached by a 20 Å cofactor. Given the fact that the *apo*-enzyme was used, this finding is not surprising, as we would assume that *apo*-PCP is in a state where it seeks interaction with a Ppan transferase at first (figure 6.13A). If it were already dedicated to the downstream C domain (figure 6.13D), the other necessary reactions postulated above would be even harder to perform. However, this is an artificial excised system where

competing interaction partners are absent. Therefore, no definite conclusion about the NRPS's general mechanics should be drawn, when it is solely based on the PCP-C-interactions seen in the bidomain structure.

An important feature observed for the first time here is the so-called linker region between the two domains. These regions have lengths between 10 and 25 amino acids, roughly, and are more or less arbitrarily defined by either the relative distance to conserved residues in the flanking domains or their terminal secondary structures. Linker regions are not conserved in sequence and are believed to merely act as natural spacers that keep the *in cis* acting catalytic entities covalently attached and in the appropriate order. The 18 amino acid linker observed here lies along the surface of the C domain. It does not show any secondary structural motifs itself, yet it undergoes weak, mostly hydrophobic interactions with side chains of the C domain's helix 5. For the constructive interaction between PCP_{TycC5} and C_{TycC6}, the flexibility of the linker region could on the other hand be a mechanistic feature that allows the PCP domain to detach from the C domain, acquire a different fold and then re-attach. In the H-form, the position of serine 43 is dramatically altered compared to the A/H: Helix 2 loses one winding at its N-terminal end, and serine 43 is now part of a more flexible loop region. So far, no structure of the interaction between a *holo*-PCP and its downstream C domain is known, but it may be assumed that one would find the carrier protein in its H-state (and possibly a different topology of the linker region). Recent studies aiming at the identification of carrier protein residues involved in inter-domain communication [Zhou 2006] have shown that recognition is dependent on specific side chain interactions. However, no conclusions about the carrier protein domain's fold can be drawn in this context, and models that are experimentally not supported would be highly speculative.

Besides the expected flexibility of PCP domains and the linker regions, it is noteworthy that the other nonribosomal domain types might in turn also undergo structural alterations during the synthetic process. This possibility is supported by the finding that the two CAT-like subdomains of condensation domains are off-rotated by ~10° when comparing the VibH structure with PCP-$C_{TycC5-6}$. Furthermore, A and TE domains are known to have

subdomains and lid regions, respectively, which might also be involved in the inter-domain communication process.

In summary, our current state of knowledge does not enable us to fully understand how NRPSs work, but it has become evident that the delicate sequence of single reactions is dependent on a dynamic structural interplay of the domains.

With the recently published structures of the yeast and fungal fatty acid synthetases (FASs) [Leibundgut 2007, Jenni 2007], a complete view of similar machinery is possible. Both FASs and NRPSs use Ppan-modified carrier proteins for the covalent tethering of reaction intermediates. With the given similarities in function, sequence, and fold between ACPs and PCPs, a closer look at FASs can be worthwhile to gain further understanding of the NRPS.

The yeast FAS is a heterododecameric complex with two large reaction chambers each carrying three ACP (18 kDa) units. Interestingly, the ACPs are anchored by two flexible peptide linkers (~40 amino acids in length), confining them to their subsection of the chamber – yet theoretically allowing them to re-orientate to interact with multiple reaction partners. The situational snapshot taken by this crystallization experiment shows the ACP in contact with a ketosynthase (KS) unit by polar interactions (Figure 6.14).

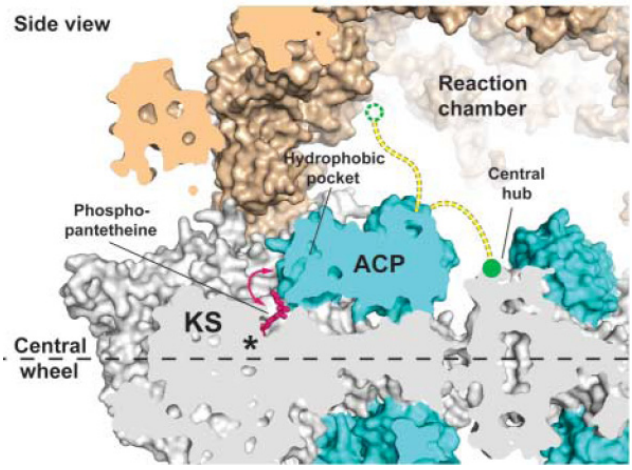


Figure 6.14: An ACP (blue) within the yeast FAS reaction chamber interacting with a KS unit (grey). The Ppan cofactor (purple) is stretched towards the active site of the KS. The two linker peptide sequences are shown in yellow and their corresponding anchor sites are indicated by green circles [Leibundgut 2007].

Even though no folding dynamics can be observed in the FAS structures, the presence and location of two flexible linker regions suggest substantial mobility of the ACP. Analogously, the PCP domains and their linkers in NRPSs are half the size/length, roughly, and it could be suspected that a similar relative amount of space is available for

PCP domain movement in NRPSs. It remains to be discovered, how much room there is for mobility of each PCP within a complete, properly folded NRPS – an information that would best be acquired from crystal structures of larger nonribosomal synthetase fragments.

7. References

Balibar, C.J., Vaillancourt, F.H., Walsh, C.T. 2005. Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem Biol*. 12(11): 1189-200.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., Steitz, T.A. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science*. 289(5481): 905-20.

Bashford, D., Karplus, M. 1990. pKa of ionizable groups in proteins: Atomic detail from a continuum electrostatic model. *Biochemistry.* 29: 10219-10225.

Belshaw, P.J., Walsh, C.T., Stachelhaus, T. 1999. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science*. 284(5413):486-9.

Bender, C.L., Alarcon-Chaidez, F., Gross, D.C. 1999. Pseudomonas syringae phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiol Mol Biol Rev*. 63(2): 266-92.

Bergendahl, V., Linne, U., Marahiel, M.A. 2002. Mutational analysis of the C-domain in nonribosomal peptide synthesis. *Eur J Biochem*. 269(2): 620-9.

Beringer, M., Rodnina, M.V. 2007. The ribosomal peptidyl transferase. *Mol Cell*. 26(3):311-21.

Billich, A., Zocher, R. 1987. N-Methyltransferase function of the multifunctional enzyme enniatin synthetases. *Biochemistry.* 26(25): 8417-8423.

Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*. 72: 248-54.

Bu. X., Wu, X., Xie, G., Guo, Z. 2002. Synthesis of Tyrocidine A and its Analogues by Spontaneous Cyclization in aqueous solution. *Org Lett.* 4(17): 2893-2895.

Buchenau, B., Thauer, R.K. 2004. Tetrahydrofolate-specific enzymes in Methanosarcina barkeri and growth dependence of this methanogenic archaeon on folic acid or p-aminobenzoic acid. *Arch Microbiol.* 182: 313-325.

Castro, B., Dormoy, J.R.,, Dourtoglou, B., Evin,B., Selve, C., Ziegler, J.C. 1976. Peptide Coupling Reagents VI1. A Novel, Cheaper Preparation of Benzotriazolyloxytris-[dimethylamino]-phosphonium Hexafluorophosphate (BOP Reagent). *Synthesis* 1976: 751-752.

Chen, H., O'Connor, S., Cane, D.E., Walsh, C.T. 2001. Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. *Chem Biol*. 8(9):899-912.

Clugston, S.L., Sieber S.A., Marahiel M.A., Walsh C.T. 2004. Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C5 domain of tyrocidine synthetase is a (D)C(L) catalyst. *Biochemistry* 42(41):12095-104.

Ehmann, D.E., Trauger, J.W., Stachelhaus, T., Walsh, C.T. 2000. Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chem Biol*. 7(10): 765-72.

Eppelmann, K., Doekel, S., Marahiel, M.A. 2001. Engineered biosynthesis of the peptide antibiotic bacitracin in the surrogate host Bacillus subtilis. *J Biol Chem*. 276(37): 34824-31.

Eppelmann, K., Stachelhaus,T., Marahiel, M.A. 2002. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry*. 41(30): 9718-26.

Finking, R., Marahiel, M.A. 2004. Biosynthesis of nonribosomal peptides 1. *Annu Rev Microbiol*. 58: 453-88.

Fischbach, M.A., Walsh, C.T. 2006. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev*. 106(8):3468-96.

Fortin, P.D., Walsh, C.T., Magarvey, N.A. 2007. A transglutaminase homologue as a condensation catalyst in antibiotic assembly lines. *Nature*. 448(7155): 824-7.

Gehring, A.M., Mori, I., Walsh, C.T. 1998. Reconstitution and characterization of the Escherichia coli enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry*. 37(8):2648-59.

Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., Onufriev, A. 2005. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucl Acids Res.* 33: W368-W371.

Griffin, M., Casadio, R., Bergamini, C.M. 2002. Transglutaminases: nature's biological glues. *Biochem J*. 368(Pt 2): 377-96.

Hahn, M., Stachelhaus, T. 2004. Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proc Natl Acad Sci* 101(44):15585-90.

Hall, H.K. Jr. 1957. *J. Am. Chem. Soc*. 79: 5441.

Hotchkiss, R.D., Dubois, R.J. 1940. Fractionation of bactericidal agents from cultures of a soil bacillus. *J Biol Chem.* 132: 791-792.

Hsiao, Y. S., Jogl, G., Tong, L. 2004. Structural and biochemical studies of the substrate selectivity of carnitine acetyltransferase. *J Biol Chem.* 279: 31584-31589.

Hubbard, B.K., Walsh, C.T. 2003. Vancomycin assembly: nature's way. *Angew Chem Int Ed Engl*. 42(7): 730-65.

Jenni, S., Leibundgut, M., Boehringer, D., Frick, C., Mikolásek, B., Ban, N. 2007. Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science*. 316(5822): 254-61.

Keating, T.A., Marshall, C.G., Walsh, C.T., Keating, A.E. 2002. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat Struct Biol*. 9(7): 522-6.

Keller, U., Schauwecker, F. 2003. Combinatorial biosynthesis of non-ribosomal peptides. *Comb Chem High Throughput Screen*. 6(6): 527-40.

Kessler, N., Schuhmann, H., Morneweg, S., Linne, U., Marahiel, M.A. 2004. The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. *J Biol Chem*. 279(9):7413-9.

Koglin, A., Mofid, M.R., Lohr, F., Schafer, B., Rogov, V.V., Blum, M.M., Mittag, T., Marahiel, M.A., Bernhard, F., Dotsch, V. 2006. Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science*. 312(5771): 273-6.

Kohli, R.M., Trauger, J.W., Schwarzer, D., Marahiel, M.A., Walsh, C.T. 2001. Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. *Biochemistry*. 40(24): 7099-108.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 227(5259): 680-5.

Lautru, S., Deeth, R.J., Bailey, L.M., Challis, G.L. 2005. Discovery of a new peptide natural product by Streptomyces coelicolor genome mining. *Nat Chem Biol*. 1(5): 265-9.

Leibundgut, M., Jenni, S., Frick, C., Ban, N. 2007. Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science*. 316(5822): 288-90.

Luo, L., Burkart, M.D., Stachelhaus, T., Walsh, C.T. 2001. Substrate recognition and selection by the initiation module PheATE of gramicidin S synthetase. *J Am Chem Soc*.123(45): 11208-18.

May, J.J., Kessler, N., Marahiel, M.A., Stubbs, M.T. 2002. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc Natl Acad Sci U S A*. 99(19): 12120-5.

Mootz, H.D., Kessler, N., Linne, U., Eppelmann, K., Schwarzer, D., Marahiel, M.A. 2002. Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes. *J Am Chem Soc*. 124(37): 10980-1.

Morikis, D., Elcock, A.H., Jennings, P.A., McCammon, J.A. 2001. Proton transfer dynamics of GART: The pH-dependent catalytic mechanism examined by electrostatic calculations. *Protein Sci*. 10(11): 2379-92.

Pedersen, L.C., Yee, V.C., Bishop, P.D., Le Trong, I., Teller, D.C., Stenkamp, R.E. 1994. Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules. *Protein Sci*. 3(7): 1131-5.

Quadri, L.E., Sello, J., Keating, T.A., Weinreb, P.H., Walsh, C.T. 1998. Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem Biol*. 5(11): 631-45.

Roche, E.D., Walsh, C.T. 2003. Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis. *Biochemistry*. 2003. 42(5): 1334-44.

Rouhiainen, L., Paulin, L., Suomalainen, S., Hyytiäinen, H., Buikema, W., Haselkorn, R., Sivonen, K. 2000. Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in Anabaena strain 90. *Mol Microbiol*. 37(1): 156-67.

Sambrook, J., Fritsch, E.F., Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual. *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.*

Samel, S.A., Schoenafinger, G., Knappe, T.A., Marahiel, M.A., Essen, L.O. 2007. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* 15(7): 781-92.

Sarges, R., Witkop, B. 1965. The structure of valine- and isoleucine-gramicidin A. *J Am Chem Soc.* 87: 2011-2020.

Sato, N.S., Hirabayashi, N., Agmon, I., Yonath, A., Suzuki, T. 2006. Comprehensive genetic selection revealed essential bases in the peptidyl-transferase center. *Proc Natl Acad Sci U S A*. 103(42): 15386-91.

Schauwecker, F., Pfennig, F., Grammel, N., Keller, U. 2000. Construction and in vitro analysis of a new bi-modular polypeptide synthetase for synthesis of N-methylated acyl peptides. *Chem Biol*. 7(4): 287-97.

Schmeing, T.M., Huang, K.S., Strobel, S.A., Steitz, T.A. 2005. An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature*. 438: 520-524.

Schmitt, E., Panvert, M., Blanquet, S., Mechulam, Y. 1998. Crystal structure of methionyl-tRNAfMet transformylase complexed with the initiator formyl-methionyl-tRNAfMet. *EMBO J.* 17(23): 6819-26.

Schoenafinger, G. 2003. Charakterisierung von internen Kondensationsdomänen der nicht-ribosomalen Peptidsynthetase des Tyrocidins: Untersuchungen zur Akzeptanz von Substraten in cis und in trans. *Diploma Thesis, Fachbereich Biochemie, Philipps-Universität Marburg, Germany*.

Schoenafinger, G., Schracke, N., Linne, U., Marahiel, M.A. 2006. Formylation domain: an essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin. *J Am Chem Soc*. 128(23): 7406-7.

Schracke, N. 2005a, Die molekulare Logik der nichtribosomalen Peptidsynthetasen: Identifizierung und biochemische Charakterisierung der Biosynthesegene für Gramicidin A. *PhD Thesis, Fachbereich Biochemie, Philipps-Universität Marburg, Germany.*

Schracke, N., Linne, U., Mahlert, C., Marahiel, M.A. 2005. Synthesis of linear gramicidin requires the cooperation of two independent reductases. *Biochemistry*. 44(23):8507-13.

Sellergren, B., Karmalkar, R.N., Shea, K.J. 2000. Enantioselective ester hydrolysis catalyzed by imprinted polymers. *J Org Chem*. 65(13): 4009-27.

Shim, J.H., Benkovic, S.J. 1998. Evaluation of the kinetic mechanism of Escherichia coli glycinamide ribonucleotide transformylase. *Biochemistry.* 37: 8776-8782.

Sieber, S.A., Walsh, C.T., Marahiel, M.A. 2003. Loading peptidyl-coenzyme A onto peptidyl carrier proteins: a novel approach in characterizing macrocyclization by thioesterase domains. *J Am Chem Soc*. 125(36): 10862-6.

Stachelhaus, T., Mootz, H.D., Bergendahl, V., Marahiel, M.A. 1998. Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J Biol Chem*. 273(35): 22773-81.

Stachelhaus, T., Mootz, H.D., Marahiel, M.A. 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol*. 6(8): 493-505.

Stachelhaus, T., Schneider, A., Marahiel, M.A. 1995. Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science*. 269(5220):69-72.

Stein, D.B., Linne, U., Hahn, M., Marahiel, M.A. 2006. Impact of epimerization domains on the intermodular transfer of enzyme-bound intermediates in nonribosomal peptide synthesis. *Chembiochem*. 7(11): 1807-14.

Townsley, L.E., Tucker, W.A., Sham, S., Hinton, J.F. 2001. Structures of gramicidins A, B, and C incorporated into sodium dodecyl sulfate micelles. *Biochemistry*. 40(39):11676-86.

Velkov T, Lawen A. 2003. Mapping and molecular modeling of S-adenosyl-L-methionine binding sites in N-methyltransferase domains of the multifunctional polypeptide cyclosporin synthetase. *J Biol Chem*. 278(2): 1137-48.

Walsh, C.T. 2004. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science*. 303(5665): 1805-10.

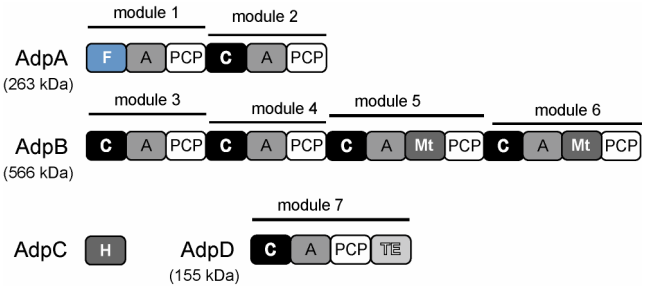
Weber, T., Baumgartner, R., Renner, C., Marahiel, M.A., Holak, T.A. 2000. Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure*. 8(4): 407-18.

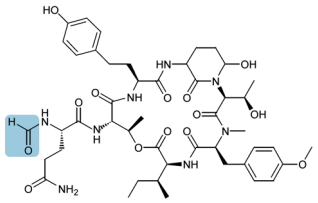
Williams, G.J., Breazeale, S.D., Raetz, C.R., Naismith, J.H. 2005. Structure and function of both domains of ArnA, a dual function decarboxylase and a formyltransferase, involved in 4-amino-4-deoxy-L-arabinose biosynthesis. *J Biol Chem*. 280(24): 23000-8.

Zhou, Z., Lai, J.R., Walsh, C.T. 2006. Interdomain communication between the thiolation and thioesterase domains of EntF explored by combinatorial mutagenesis and selection. *Chem Biol*. 13(8): 869-79.

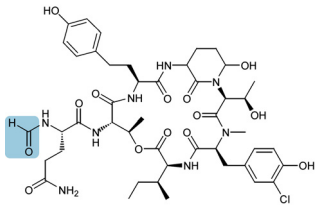
8. Appendix

Biosynthetic Enzymes Involved in the Production of Anabaenopeptilides:





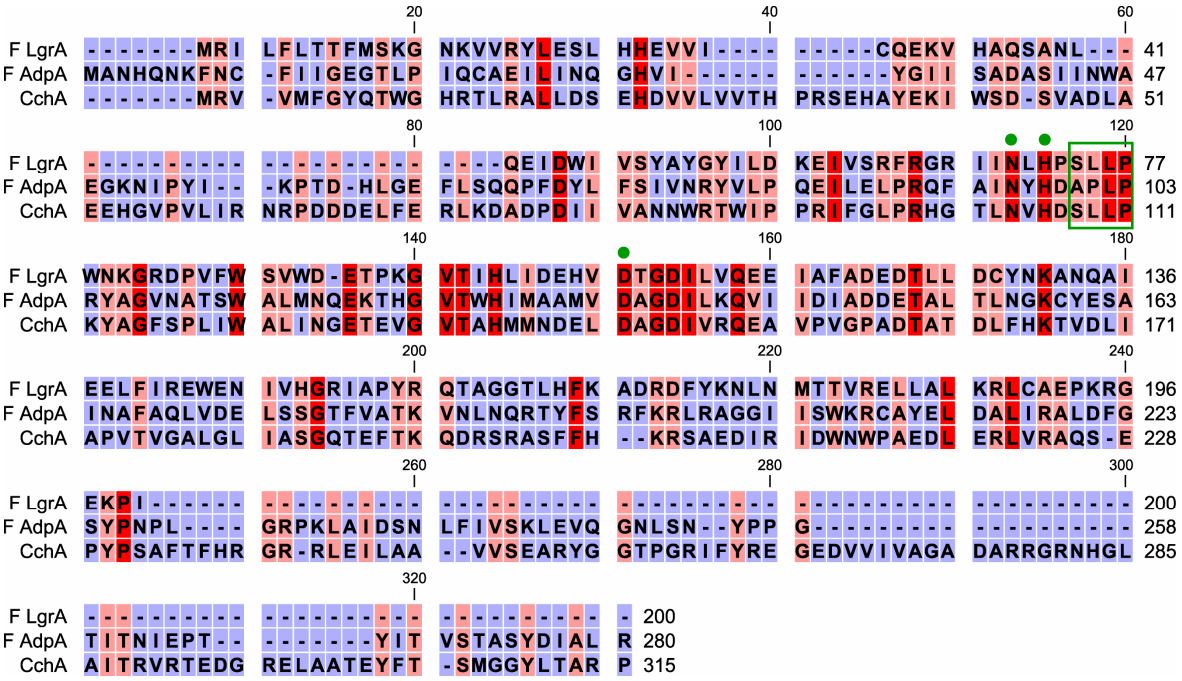
Anabaenopeptilide 90 A



Anabaenopeptilide 90 B

The 263 kDa synthetase AdpA consists of two modules (6 domains). The N-terminal domain is a putative formyltransferase. AdpB (566 kDa) comprises four modules, the third and fourth of which feature N-methyltransferase (Mt) domains situated between the A and PCP domains. AdpC has been suggested by sequence comparisons [Rouhiainen 2000] to be halogenase, and AdpD is a one-module synthetase with a terminal thioesterase (TE) domain. From these putative synthetases alone, the composition of the products cannot be explained thoroughly. It appears that the primary sequence of the depsipeptides is Gln1-Thr2-homoTyr3-aminohydroxypiperidone4-Thr5-Tyr6-Ile7. Consequently, the A domains' specificies should be directed towards these building blocks. The thioesterase-catalyzed macrolactonization should involve the C-terminal Ile7 carboxyl group and the side chain hydroxyl group of Thr2. The Mt domain of module 6 obviously N-methylates Tyr6, but the origin of the aminodydroxypiperidone and the function of the second Mt domain in module 5 remain opaque. Possibly, the either Mt domain is responsible for the methylation of the side chain of Tyr6 in the 90A product. F_{AdpA} is believed to be responsible for the N_{α}-formylation of Gln1.

Sequence Alignment of Three (Putative) Formyltransferases from NRPSs:



In this alignment, three formyltransferases associated with nonribosomal peptide synthesis are presented. F_{LgrA} is characterized and discussed in this work. F_{AdpA} is the N-terminal domain of AdpA which is believed to analogously formylate the first building block *in cis*. CchA, however is genetically associated with the putative NRPS responsible for the production of coelichelin. Coelichelin carries two N_{δ}-formylated and hydroxylated ornithine residues at the N- and C-terminus (figure 1.2) [Lautru 2005]. To date, it is unclear how these modifications are realized in nature. The residues involved in the formyltransferase activity are strictly conserved (green dots), and the characteristic "SLLP" motif is highlighted by a green box.

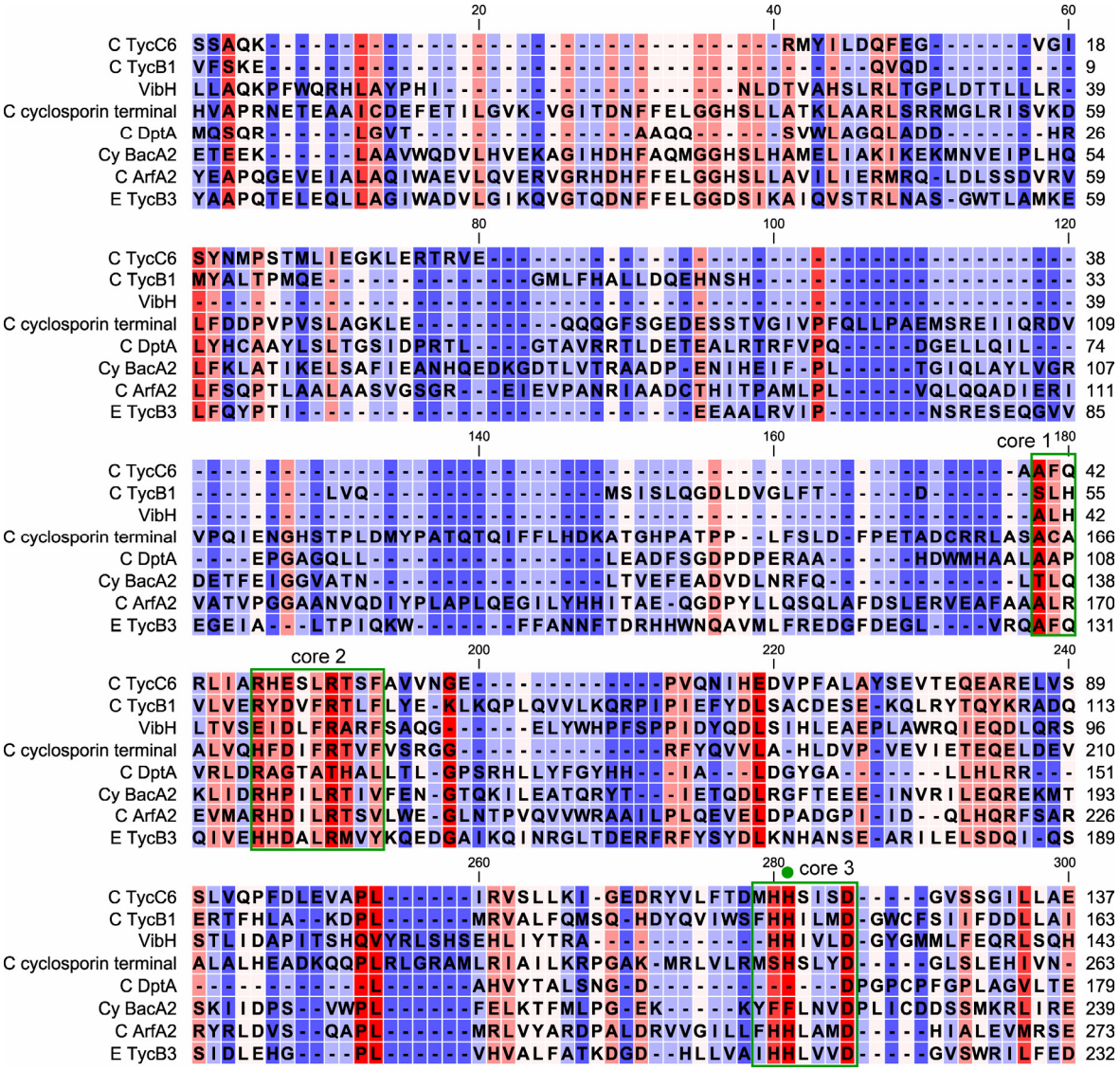
Core Motifs of Condensation Domains:

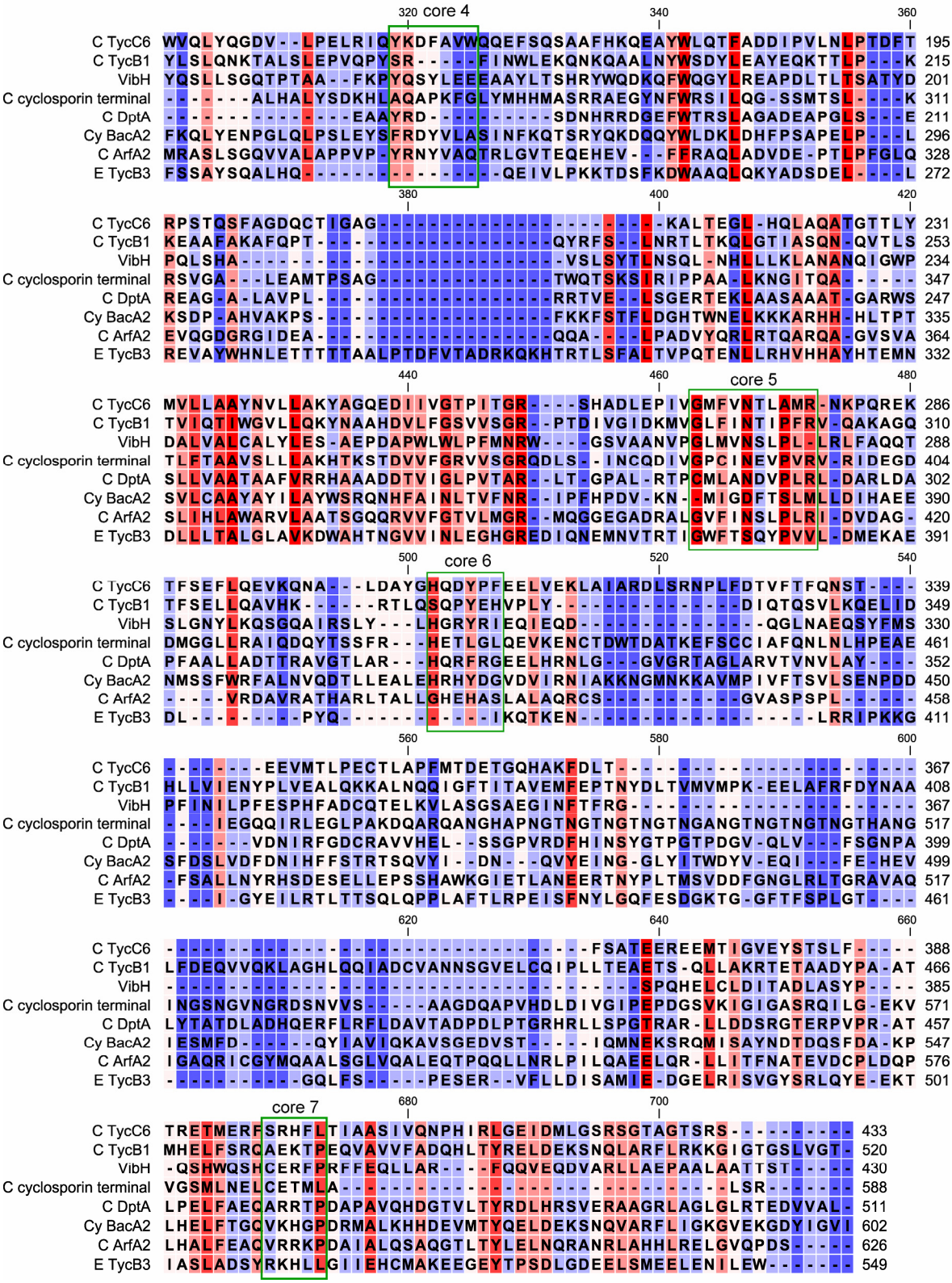
Core 1	SxAQxR(LM)(WY)xL
Core 2	RHExLRTxF
Core 3	MHHxxxDG(WV)S
Core 4	YxD(FY)AVW
Core 5	(IV)GxFVNT(QL)(CA)xR
Core 6	(HN)QD(YV)PFE
Core 7	RDxSRNPL

The (deputedly) catalytic histidine residue of core 3 is indicated in red. Residues named here are not strictly conserved, but represent an overall sequence tendency found among C domains which is commonly used to identify them when mining genomes.

Sequence Alignment of Condensation Domains

In chapter 1.3 the various functions of C domains have been pointed out (see figure 1.12). Here, an alignment of these C domains is presented. The core motifs have been mapped by green boxes, and the second histidine residue of core motif 3 is indicated by a green dot.




core 4
core 5
core 6
core 7
C TycC6
C TycB1
VibH
C cyclosporin terminal
C DptA
Cy BacA2
C ArfA2
E TycB3

Acknowledgements

Mein besonderer Dank gilt Herrn Prof. Dr. Mohamed A. Marahiel, durch dessen eingehende wissenschaftliche und persönliche Unterstützung diese Arbeit erst möglich wurde. Seine stete Diskussionsbereitschaft und seine Bereitwilligkeit, neue Ideen finanziell und ideell zu unterstützen, rechne ich ihm hoch an. Ich möchte mich des Weiteren bei ihm für die zahlreichen Auslandsaufenthalte, die ich im Rahmen des Promotionsstudiums unternehmen durfte, bedanken.

Herrn Prof. Dr. Lars-Oliver Essen möchte ich für die vielen anregenden Fachgespräche und für die freundliche Übernahme des Zweitgutachtens dieser Arbeit danken.

Herrn Prof. Dr. Norbert Hampp und Herrn Prof. Dr. Armin Geyer danke ich dafür, dass sie sich als weitere Mitglieder der Prüfungskommision bereitwillig zur Verfügung gestellt haben.

Für die angenehme gemeinsame Zeit im Labor, bei Seminaren und Ausflügen möchte der gesamten Arbeitsgruppe Marahiel meinen Dank aussprechen. Insbesondere möchte ich meinen Laborkollegen aus „4716“ danken, mit denen ich exzellent zusammen arbeiten und konstruktiv diskutieren konnte: Dr. Thomas Dürfahrt, Dr. Robert Finking, Dr. Martin Hahn, Thomas Kurpiers, Thomas Knappe, Katharina Hoyer, Melanie Wittmann und Shashipavan Chillappagari.

Darüber hinaus möchte ich mich bei Dr. Nadine Schracke, Lars Robbel, Christoph Mahlert, Dr. Jan Grünewald, Verena Pohlmann, und Alan Tanovic für zahlreiche anregende fachliche Gespräche und deren Hilfsbereitschaft bedanken.

Dr. Uwe Linne danke ich für die technischen Hilfestellungen bei analytischen Problemen.

Für technische Assistenz gilt mein besonderer Dank Antje Schäfer, Christiane Bomm und Gabriele Schimpff-Weiland.

Bei Roswitha Roller-Müller möchte ich mich für die unermüdliche organisatorische Arbeit und ihre stete, freundliche Hilfsbereitschaft bei Alltagsproblemen bedanken.

Bei Dr. Anke Raufuß, Dr. Holger Benthien und Lars Robbel möchte ich mich für das Korrekturlesen dieser Arbeit bedanken. Anke gilt darüber hinaus mein besonderer Dank für die beständige liebevolle Unterstützung während meiner gesamten Studienzeit und den zahllosen gemeinsam verbrachten Stunden. Ihr widme ich diese Arbeit.

Stefan Samel möchte ich für die angenehme Zusammenarbeit während unseres gemeinsamen Studiums und der Promotionszeit danken.

Für ideelle und finanzielle Unterstützung danke ich der Studienstiftung des deutschen Volkes sowie dem deutschen akademischen Austauschdienst.

Bei meiner Familie und insbesondere bei meinen Eltern möchte ich mich von ganzem Herzen bedanken. Für alles.

Erklärung

Ich versichere, dass ich meine Dissertation
„Amide Bond Formation in Nonribosomal Peptide Synthesis: The Formylation and Condensation Domains"
selbständig, ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich angegebenen Quellen und Hilfen bedient habe.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

(Ort, Datum) (Unterschrift mit Vor- und Zuname)